CDAE-R: Multifunctional End-to-End Model for Brain Abnormality Images Classification and Denoising

Zezhou Wang

School of Computing, The Australian National University, Canberra, 2601, Australia
u7439262@anu.edu.au

Abstract. Traditionally, medical image classification and denoising tasks are conducted and evaluated separately, which may waste computational resources and incur excessive expenses. Besides, the features extracted by different models cannot be shared and utilized effectively. Therefore, an end-to-end multimodal neural network model is required and of importance. For the first time, this study attempted and proposed a multimodal autoencoder-based model, Classification, and Denoising Autoencoder with Residual blocks (CDAE-R) on the dataset composed of three brain abnormalities, aneurysm, tumor, and cancer for image denoising and abnormality classification. The study implemented the basic structure of an autoencoder, consisting of an encoder and an interconnected decoder, as the model's foundation. The concept of residual learning was incorporated into CDAE-R in the form of residual blocks, enabling the preservation of essential features. On top of the traditional denoising autoencoder, the study employed and connected the 2D convolutional dense layers and pooling layers after the latent space at the end of the autoencoder component for classification. CDAE-R was then trained and evaluated on joint loss of classification and denoising with equal weight coefficient. The experiment results achieved 100% classification accuracy, competitive denoising, and reconstruction performance on the test dataset, indicating the effectiveness and feasibility of CDAE-R.

Keywords: Autoencoder, Brain Abnormality Images, Residual Learning.

1 Introduction

The brain is the most significant organ in the human body, but it is fragile and extremely sensitive to some brain abnormalities, e.g., aneurysms, tumors, and cancer. They may affect the health of the human brain and even threaten human life. A brain aneurysm is an abnormal bulge in the wall of a brain blood vessel, such as a brain artery, at risk of rupturing and bleeding. A Brain tumor refers to a group of abnormal cells in the brain due to uncontrolled cell proliferation caused by genetic mutations. Brain tumors can be divided into two types: benign and malignant. The benign brain tumor is not cancerous and has no risk of metastasis and spread. Conversely, the malignant brain tumor, also typically called brain cancer, is a cancerous tumor that is at risk of spreading between different organs and is difficult to cure. In modern medicine, various medical imaging techniques have been developed for scanning the brain and detecting abnormalities due

Y. Wang (ed.), *Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024)*, Advances in Computer Science Research 115,
https://doi.org/10.2991/978-94-6463-540-9_93

to the invisibility of the brain's inner conditions to ensure the health of the human brain. However, affected by the structural isomorphism of different brain abnormalities and factors like electrical interference, it is challenging to always make accurate diagnoses for medical practitioners in large-scale situations.

Machine learning techniques have been considered recently for brain disease image classification due to their superior feature extraction capability and prediction performance in many tasks [1-4]. Convolutional Neural Network (CNN) is particularly suitable for tasks related to image processing and classification. Ayadi et al. constructed a deep CNN model for classifying brain tumors and achieved the ideal outcome [5]. Sejuti et al. combined the CNN and Support Vector Machine (SVM) for brain tumor classification in their work, and the proposed CNN-SVM-based model finally provided an accuracy of 97.1% [6]. Autoencoders have also been utilized in this field. For instance, in their study, Vaidhya et al. implemented the Stacked Denoising Autoencoder (SDAE) for image denoising [7]. Nayak et al. adopted a deep autoencoder model for brain tumor detection. They achieved an advanced accuracy of 97% and a Receiver Operating Characteristic Curve (ROC) curve area of 99.46% compared to previous models [8]. Despite the fact that massive research has been conducted in the field of brain abnormality identification, the hybrid approach of combining the autoencoder and CNN for multiple purposes has rarely been studied in recent literature.

Therefore, the paper delves into applying autoencoders and CNN techniques in medical image denoising and classification and proposes an autoencoder-based structure for both purposes. The experiment conducted by the paper can reduce the influence of objective interferences by image denoising module, distinguish between three different brain abnormalities, aneurysm, brain tumor, and brain cancer, by classification module, and assist medical practitioners in making diagnosis.

To be more specific, the paper employs the basic autoencoder structure as the foundation of the proposed variation, Classification, and Denoising Autoencoder with Residual block (CDAE-R) and selects the Computed Tomography (CT) of the human brain as the object of research. CDAE-R incorporates the residual blocks into its encoder and decoder structure, enabling these components to "remember" and reuse the valuable features in the input images. On the other hand, CDAE-R connects the dense convolutional layers and pooling operations to the latent space obtained through the encoder for classification tasks. Eventually, CDAE-R achieved 100% classification accuracy in the testing dataset and can successfully filter out the noise in the input images and reconstruct with significant features from CT, contributing to the denoising of medical images and identification of three brain abnormalities.

2 Method

2.1 Dataset Preparation

The data adopted in this paper was collected and shared on Kaggle by the Training Data group from Abu Dhabi [9]. This dataset contains 256 JPEG images, which have been divided into three folders: "aneurysm," "cancer," and "tumor." Each image is composed of RGB channels, and the resolution is 512×512 pixels. Besides, the dataset also

contains the corresponding DCM files of each JPEG image for medical use, which is not used in the context of this study. Fig. 1 presents a sample image of the brain tumor from the dataset above.

Fig. 1. Sample tumor image in the dataset.

Five abnormal images produced during the original data generation process, three from the aneurysm folder and two from the cancer folder, were removed in this study. Then, the experiment performed a series of data preprocessing operations to prepare the dataset for subsequent processing. The preprocessing pipeline is constructed based on '**transforms**' module from Pytorch framework and organized in following sequence.

- **Random Horizontal Flip**: Horizontally flipping the image with a probability of 0.5 increases the diversity of data and prevents the model from overfitting.
- **Random Rotation**: Images are randomly rotated by up to 10 degrees for both directions, introducing variability to the data.
- **Tensor Conversion**: Images are converted into tensors to keep in line with the Pytorch framework.
- **Gaussian Noise Addition**: Gaussian noise with a mean of 0 and a standard deviation of 0.15 is added to the images. This step facilitates the training of the denoising component of the model while introducing robustness to the data.
- **Salt-and-Pepper Noise Addition**: Salt-and-Pepper noise is added with a probability of 0.01 to introduce further variability to the data.
- **Normalization**: The study employs the mean and standard deviation of the ImageNet data [10] ([0.485, 0.456, 0.406]) and ([0.229, 0.224, 0.225]) as mean and standard deviation, respectively, for data normalization. Expecting that the utilization of pre-trained parameters from ImageNet can enhance the effectiveness of transfer learning.

A sample after the preprocessing pipeline is provided in Fig. 2 below.

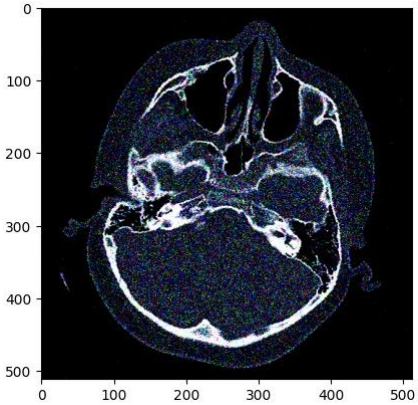

Fig. 2. A sample after the preprocessing pipeline .

2.2 CDAE-R

The Introduction of Autoencoder. Autoencoder is an unsupervised learning technique for constructing a deep neural network consisting of two interconnected components: an encoder and a decoder. The encoder can extract the essential features from input data and encapsulate them in a lower-dimensional feature representation, while the decoder aims to accurately reconstruct the input by generating details from this representation. This feature representation is called latent space in the context of the autoencoder, which contains the core information of the input data. The structural superiority enables the autoencoder to derive useful information and filter noise in the input data, making the autoencoder suitable for data compression and denoising in the practice. Autoencoder is widely applied and has many variations, such as the Denoising Autoencoder (DAE) used in this study.

The Proposed CDAE-R Model. The study proposes the Classification and Denoising Autoencoder with Residual block (CDAE-R) for brain image denoising and abnormality identification. CDAE-R extends the basic structure of the denoising autoencoder and connects the dense layers to the latent space for the classification task. The concept of residual learning proposed in ResNet [11] is incorporated into the encoder and dense layer in the CDAE-R. The architecture of the proposed CDAE-R model is provided in Fig. 3 below.

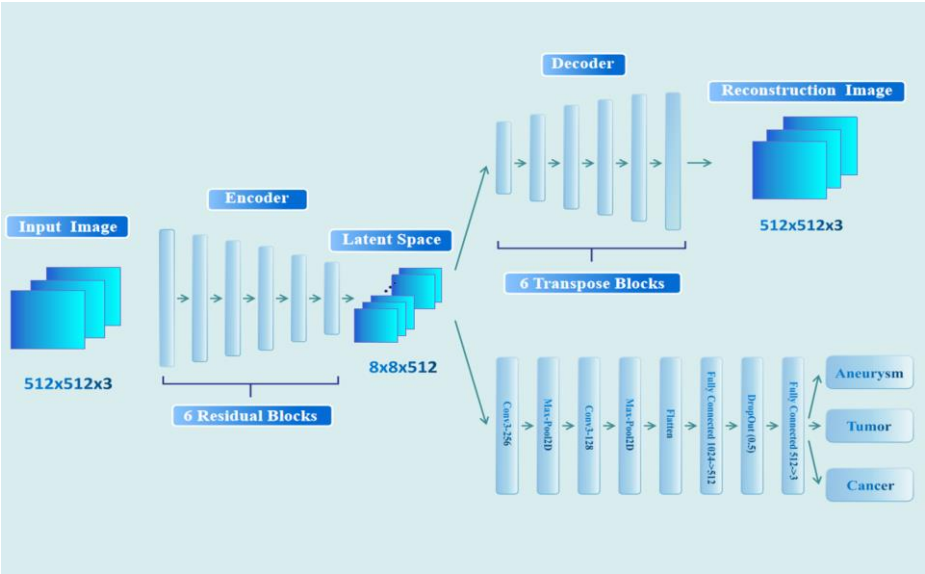

Fig. 3. The architecture of CDAE-R .

CDAE-R passes the input data to the encoder. The encoder consists of six interconnected residual blocks. The details of the residual block, along with the transpose block, are provided in Fig. 4 below. Each residual block comprises two 2D convolutional layers that use 3×3 kernels, with a stride of 2 for downsampling at the first layer and a padding of 1 for both layers to preserve spatial dimensions. The two layers process tensors through 2D batch normalization and Rectified Linear Unit (ReLU) activation function. From the second residual block, the downsampled original data from the previous block will directly integrate into the current block for training. The encoder component will transform the original input of (3,512,512) into the latent space of (512,8,8). Then, the latent space will be utilized for denoising and classification.

The decoder is the encoder's transpose, consisting of six interconnected transpose blocks for upsampling and residual connection. Like a residual block, a transpose block consists of two inverse layers of 2D convolution, which generate features and increase the size. Original information from previous blocks will be compressed and added to the following blocks. The decoder shares the same settings of stride as well as padding with the encoder and processes features with 2D batch normalization and ReLU activation function. After passing through the encoder, the latent space of (512,8,8) will be reconstructed to the original input size of (3,512,512).

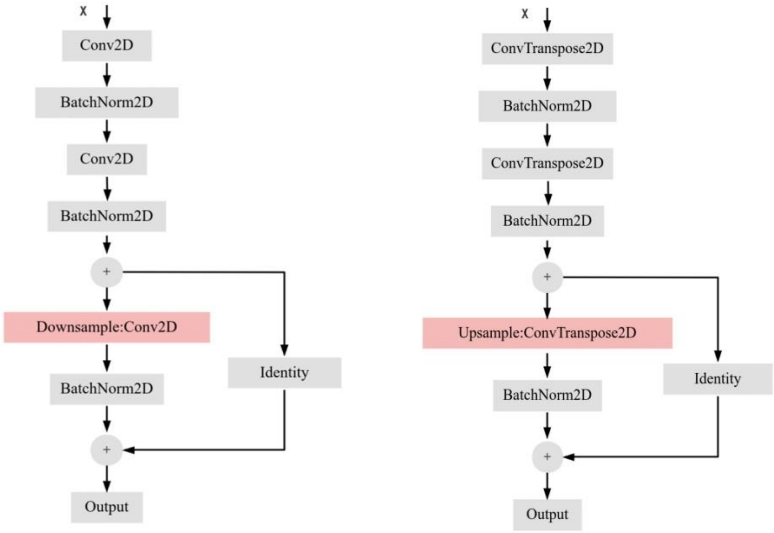

Fig. 4. Details of The Residual Block (Left) and The Transpose Block (Right) .

Simultaneously, the latent space will be utilized as the input for the classification component, consisting of convolutional layers and pooling layers, a fully connected layer followed by a dropout layer, and an output layer. Two 2D convolutional layers implement the 3×3 kernels and ReLU activation function for processing. Each layer is followed by a max pooling layer with a kernel size of 2×2 and a stride of 2. As shown in Fig. 3, the latent space of size (512,8,8) will be further compressed into size (128,2,2) for the subsequent layers. The classification component will then flatten the tensor to a 1-D format and feed it into the fully connected layer for feature learning, outputting 512 features. These features will be fed into the output layer to predict three brain abnormalities after a dropout layer at a rate of 50%.

Implementation details. The study was entirely based on the Pytorch framework. During data processing, the experiment loaded the dataset with a batch size of 32 and randomly selected training data with a probability of 80%. CDAE-R was trained by Adam optimizer with a learning rate of 0.001 and converged around the 35^{th} epoch. Mean Square Error (MSE) and cross-entropy loss were employed as the loss functions for reconstructing and classification error, respectively. Adam was then optimized with joint loss of MSE and cross-entropy loss for denoising as well as construction and classification, respectively. The weight coefficients of the two loss functions were set to be the same. Eventually, the study evaluated the performance of CDAE-R with the effect of reconstruction and classification accuracy on the remaining 20% of data as the testing dataset.

3 Results and Discussion

3.1 Denoising and Reconstruction Performance

In the denoising module, CDAE-R achieved significant denoising performance and favorable reconstruction performance on the testing dataset. The sample results are shown in the Fig. 5:

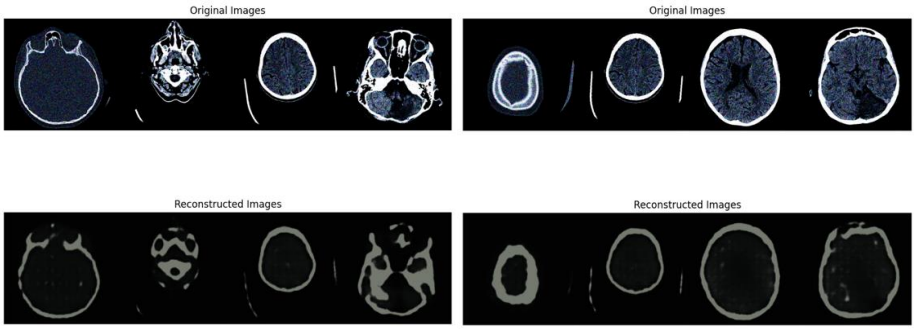

Fig. 5. Reconstruction performance of CDAE-R's on sample testing data .

It is observable that the reconstructed images successfully removed the previously added Gaussian and Salt-and-Pepper noise. Meanwhile, the reconstructed images retained the main features and most of the prominent contours in the original CT images.

The denoising results of CDAE-R indicate that the autoencoder is still a promising structure for image denoising due to its unique bottleneck mechanism. The latent space of the bottleneck will automatically extract the essential features of input images through model training and facilitate the reconstruction. By incorporating residual blocks and dropout layers, CDAE-R is less likely to get stuck in the local minimal and be impacted by gradient exploding or vanishing than vanilla DAE with an unstable training process. During experiments, CDAE-R converged stably and swiftly at approximately the 35^{th} epoch and achieved enhanced results based on residual reusing.

However, it is worth noticing that the reconstruction results are imperfect, and some details of the original images were still lost in this process. This may be attributed to the convolutional layers and the large convolution size implemented in the model. The encoder component of CDAE-R consists of 6 residual blocks. Each block has two 2D convolutional layers with a 3×3 kernel for convolution. Besides, the first convolutional layer employed a stride value of 2 for data downsampling. These techniques are practical and suitable for processing images. However, since the kernels averaged the values of 9 pixels and the stride value of 2 skipped some pixel points, some details in the original images are inevitably lost during the continuous convolution and downsampling operations.

3.2 Classification Performance

From a classification perspective, CDAE-R demonstrated exceptional performance. As shown in Table 1 below, CDAE-R achieved a 100% accuracy rate on both the training and testing datasets, indicating that the model can perfectly distinguish between three different categories, which represent aneurysm, tumor, and cancer, respectively.

Table 1. The Classification Performance of CDAE-R.

Metrics	Aneurysm	Tumor	Cancer	Overall
Training Acc.	100%	100%	100%	100%
Testing Acc.	100%	100%	100%	100%

Differing from complicated model architectures of previous approaches, CDAE-R employed the features obtained by the autoencoder's latent space as the input for classification and produced improved classification performance.

CDAE-R was trained with joint loss composed of two different loss functions, cross-entropy loss and MSE loss, for separate classification and denoising modules. Both of the loss functions can influence the final latent space. The joint training of these two modules converged stably and yielded favorable outcomes on both tasks, demonstrating that the classification, denoising, and reconstruction processes have similar expectations of what the latent space is like. The similar expectation means that the classification and denoising are consistent in some aspects and may further indicate that the latent space has the capability to be competent in multimodal task training.

In this study's experiment, the final weight coefficients of two loss functions are set to be the same. Performances of different weight coefficient combinations have also been explored. If the classification component is more emphasized, the performances of both modules will worsen, and if the denoising component is more emphasized, classification accuracy will decrease significantly. Therefore, CDAE-R needs to make a trade-off between these two parts.

4 Conclusion

This work proposed the CDAE-R based on the basic structure of DAE for medical image denoising as well as classification. The work connected dense convolutional layers along with pooling layers after the latent space of the autoencoder for the classification task. It incorporated the concept of residual learning into the model for the reuse of previous critical features. The study conducted experiments on the dataset of human brain's CT images, which contains three different brain abnormalities: aneurysms, tumors, and cancer. Experiment results achieved 100% classification accuracy and remarkable denoising and reconstruction performances, demonstrating the competence of the bottleneck mechanism (the latent space) of the autoencoder in handling classification and denoising tasks simultaneously. Future work can extend the experiments to other medical images, such as MRI, to improve the model's ability when facing different data types, scales, and category quantities or integrate advanced pre-trained models, such as graphGPT, for enhanced reconstruction performance.

References

1. Liu, Y., Liu, L., Yang, L., Hao, L., Bao, Y.: Measuring distance using ultra-wideband radio technology enhanced by extreme gradient boosting decision tree (XGBoost). Automation in Construction 126, 103678 (2021).
2. Qiu, Y., Hui, Y., Zhao, P., Cai, C. H., Dai, B., Dou, J., Bhattacharya, S., Yu, J.: A novel image expression-driven modeling strategy for coke quality prediction in the smart cokemaking process. Energy 294, 130866 (2024).
3. Zhao, F., Yu, F., Trull, T., Shang, Y.: A new method using LLMs for keypoints generation in qualitative data analysis. In 2023 IEEE Conference on Artificial Intelligence (CAI), pp. 333-334. IEEE (2023).
4. Liu, Y., Bao, Y.: Intelligent monitoring of spatially-distributed cracks using distributed fiber optic sensors assisted by deep learning. Measurement 220, 113418 (2023).
5. Ayadi, W., Elhamzi, W., Charfi, I., Atri, M.: Deep CNN for brain tumor classification. Neural Processing Letters 53, 671-700 (2021).
6. Sejuti, Z. A., Islam, M. S.: An efficient method to classify brain tumor using CNN and SVM. In 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), pp. 644-648. IEEE (2021).
7. Vaidhya, K., Thirunavukkarasu, S., Alex, V., Krishnamurthi, G.: Multi-modal brain tumor segmentation using stacked denoising autoencoders. In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: First International Workshop, Brainles 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Revised Selected Papers 1, pp. 181-194. Springer International Publishing (2016).
8. Nayak, D. R., Padhy, N., Mallick, P. K., Singh, A.: A deep autoencoder approach for detection of brain tumor images. Computers and Electrical Engineering 102, 108238 (2022).
9. Kaggle Website (dataset): Computed Tomography (CT) of the Brain. Available at https://www.kaggle.com/datasets/trainingdatapro/computed-tomography-ct-of-the-brain (Accessed on [1 Apr, 2024]).
10. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255. IEEE (2009).
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778 (2016).