Enhancing Tumor Classification In MRI Images with Ghosting Artifacts

Mengde Liu
Department of Biomedical Engineering
Duke University
ml720@duke.edu

Emma Nisbet
Department of Electrical and Computer Engineering
Duke University
ekn18@duke.edu

Talya Jeter
Department of Biomedical Engineering
Duke University
tj95@duke.edu

Abstract

This project fine-tunes a ResNet-50 model for multi-classification that accommodates ghosting artifacts for MRI brain tumor diagnostics. We incorporated a physical layer of our model that generated ghosting in the horizontal, vertical, and diagonal directions. We then took a weighted sum of each ghosted direction with the original image. With this model, we were able to classify between glioma, meningioma, pituitary tumor, and no tumor groups with an accuracy of 91.69%. We found the ghosting direction that had the most important weight for classification accuracy was horizontal ghosting and vertical ghosting was the least important. Only 0.77% of tumor images were misclassified as no tumor with our model.

1 Introduction

According to the National Cancer Institute, the age-adjusted rate of new brain cancer cases is around 6.2 per 100,000 people per year [SEER, 2024]. The timing of diagnosis has a strong influence on a patient's 5-year survival rate where brain tumors caught at the earliest stage have a survival rate of 35.5% whereas the survival rate is only 21.3% for the second stage. Additionally, the prognosis will vary depending on tumor type where the survival rate for Stage 4 gliomas is merely 6% for ages 55-64 [American Cancer Society, 2024]. The decreasing survival rate depending on stage indicates a significant need for accurate, early diagnostics for brain cancers.

Magnetic Resonance Imaging (MRI) serves as a non-invasive way to diagnose brain tumors with 2D or 3D images. MRI works by using a large magnetic pulse that aligns the magnetic moment hydrogen atoms in the direction of the magnet [Preston, 2016]. To measure the net magnetization of the hydrogen atoms within the patient, a short radiofrequency pulse is applied perpendicularly to the primary magnetic pulse to displace the hydrogen atoms in a new plane. Then, a receiver coil can read the current induced by the net magnetization from the hydrogen atoms. Spatial-specific information is stored by applying linear gradients across the plane so the behavior of hydrogen atoms is unique to their position along the gradient. Next, the Fourier Transform stores the spatial frequency information in the frequency domain, or k-space. The image is reconstructed with gray-scale intensities based on the inverse Fourier Transform of the pixels in k-space [Gallagher et al., 2008].

MRI images are subject to artifacts during imaging acquisition including motion artifacts and hardware-induced phase shifts. These artifacts can reduce the quality of images making it difficult for radiologists to identify tumors. One artifact of concern is ghosting, which corrupts sampling in k-space, creating low-intensity replicates of the original object across the image after

reconstruction[Zaitsev et al., 2015]. Typically, the motion responsible for ghosting includes periodic motion like breathing or swallowing, but can also be generated upon patient movement. Radiologists can choose a phase-encoding direction that does not interfere with these motions, however, to analyze some tumors and for some patients, this can be unavoidable. As such, there is a need to be able to readily interpret ghosted images.

Current deep-learning research on ghosting artifacts in MRI images pertains to ghosting correction. Kyathanahally and Kreis, 2018 used a Deep Convolutional Neural Network for ghost detection and used a Stacked What-Where Auto Encoder for subsequent ghost removal [Kyathanahally et al., 2018]. The ghost detection model had an accuracy of close to 100% while the ghost removal model had a lower accuracy they did not formally report. They found when combining the models, inaccuracies would propagate to the final reconstructed image. Another drawback to image reconstruction before classification is the extended computational time before a final diagnosis. Ghost correction algorithms assume that it is necessary to remove ghosting to have accurate diagnostics, which may be true for clinical radiologists, however, machine learning models could make use of hidden features in ghosted images to make more accurate classifications.

Our goal is to fine-tune a ResNet-50 model for multi-class classification on MRI images that is robust to various ghosting directions. We wish to classify between 4 categories: glioma, meningioma, pituitary tumor, and no tumor. We aim to input a weighted sum of the original image and horizontal, vertical, and diagonally ghosted images as a layer in our model. We will then retrieve the weights to identify which combination of image perturbations the machine learning model prefers to make a diagnostic. Finally, we wish to analyze the confusion matrix of our model to see which classes our model has the most difficulty distinguishing between.

2 The Brain Tumor MRI Dataset

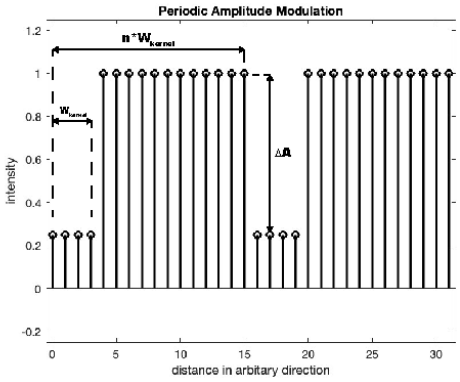
The brain tumor MRI dataset is a CC0: Public Domain dataset that contains 7023 images of human brain MRI images published on Kaggle [Nickparvar, 2021]. Our images include a mixture of T1-weighted, T2-weighted, and FLAIR MRI images as well as coronal, sagittal, and transverse slices. The images are classified and labelled as glioma, meningioma, no tumor and pituitary and split into training (N = 5712) and testing (N = 1311) data by the author of the dataset. The tumor class images are taken from the brain tumor dataset [Cheng, 2017] and the SARTAJ dataset [Bhuvaji et al., 2020] . It is worth mentioning that the SARTAJ dataset has issues in its classification of glioma images and thus the author omitted SARTAJ when creating the glioma subset. All no-tumor class images originated from the Br35H dataset [Hamada, 2020].

3 Ghosting artifact simulations

We simulated ghosting artifacts based on the mathematical description by Reeder et al. [Reeder et al., 1997]. Assuming no echo-time shifting is used and no echo misalignment occurs due to system filters, the only corruptions of k-space are periodic amplitude modulations [King et al., 1995]. Therefore, the first step of the simulation is to convert image data to k-space data via Fourier transformation. We then created a mask to apply the amplitude modulations to k-space data in an alternating pattern. The mask is initialized to be an array of the same size as the k-space image filled with all 1s. An amplitude modulation A is then applied to the mask at locations prescribed by modulation kernel width, W_{kernel}, and the period of modulation in terms of kernel width, n. In other words, given the direction of the amplitude modulation i, every first W_{kernel} pixels in a $n * W_{kernel}$ period will be assigned a new value $1 - A$ as shown in Figure 1.

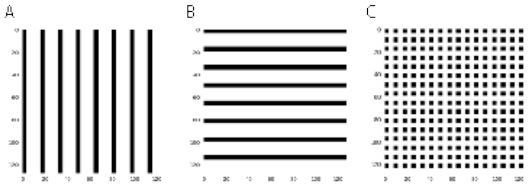
For our simulations, amplitude modulation masks for ghosting in horizontal, vertical, and diagonal directions are created separately, as shown in Figure 2. We prescribed A to be 0.75 and W_{kernel}, to be 4 and the period of modulation with respect to modulation kernel width, n, to also be 4 to create horizontal and vertical ghosting effects. To keep the amount of modulation consistent across all types of ghosting simulation, for diagonal ghosting, we changed the modulation kernel width and the period of modulation with respect to modulation kernel width to 2. After multiplying the mask with the k-space data, inverse Fourier transformation is then applied to convert the modified images back to the regular spatial domain. Each type of ghosting is applied once to an image to create one ghosted image corresponding to the original image. The three ghosted images created by the physical layer

Figure 1: Periodic Amplitude Modulations



are then added to the original image as three additional layers to transform the training and testing subsets of images to be arrays of $N * 128 * 128 * 4$, where N is the number of images in the subset.

Figure 2: (A) amplitude modulation mask to create ghosting artifacts in the horizontal direction; (B) vertical direction; (C) diagonal direction



4 Model architecture

4.1 Physical layer

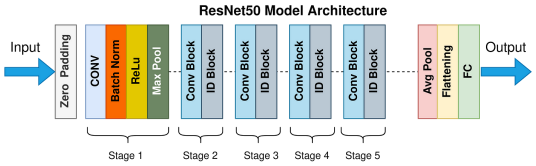
The physical layer in our model creates a weighted sum of the 4 images corresponding to an original image and outputs a single image for further training. In that case, we want to train 4 float values, one for each ghosted image (including the original, un-ghosted image) in this physical layer. Therefore, the weights are made trainable in the neural network and initialized to 1, such that at the beginning of the training every image contributed to the output of the physical layer roughly equally. A physical constraint is not enforced for the physical layer though ghosting artifacts have to be non-negative, because we prescribed all ghosting artifact simulation parameters to be positive values to create ghosted images, and it is within reason to subtract images during image processing.

4.2 Fine-tuned ResNet-50

The output from the physical layer is used as input to a pre-trained ResNet-50 model. The ResNet-50 architecture is a 50 layer variant of the Residual Network (ResNet) proposed by Kaiming He et al. in 2015. ResNet models are known for their deep learning architecture that introduce residual blocks. These blocks consist of convolutional layers, activation functions, batch normalization, and skip connections. The skip connections skip one or more layers by adding the output of a previous layer directly to the output of a deeper layer. This allows the network to learn residual mappings rather than attempting to learn the desired underlying mapping directly. This successfully finds a solution to the vanishing gradient problem, where gradients become too small as they propagate backward through many layers [He et al., 2015]. Therefore, the ResNet-50 model is a very effective deep learning model. Figure 3 shows the general structure of the ResNet-50 model.

ResNet models are commonly used for image classification and are known as state-of-the-art models for many image tasks, specifically why it was chosen to fine-tune for this project. This specific model

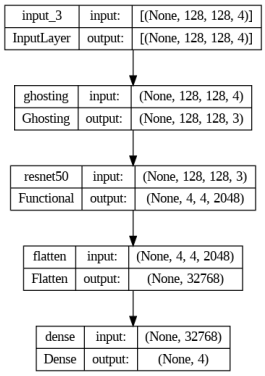
Figure 3: ResNet-50 Architecture



utilizes weights pre-trained on the ImageNet dataset, which consisted of about 14 million labeled images. Therefore, considering its deep architecture and massive pre-training dataset, it provides opportunities for transfer learning with high-performance metrics.

To fine-tune the ResNet-50 model, we added in a fully-connected layer for classification. This layer outputs logits for each of the four classes, which are unnormalized prediction scores. Combining these logits with softmax activation, the final output of the model is prediction probabilities for the four classes for each input image. The highest probability corresponds with the model's predicted class. Below is the architecture for the fine-tuned model used for this project.

Figure 4: Fine-Tuned Model Architecture



4.3 Hyper-parameters

The optimizer used throughout the project was Adam with learning rate = .001. Adam is a popular optimizer, based on adaptive estimates of lower-order moments [Kingma and Ba, 2014]. The learning rate of .001 is the standard rate for Adam and it yielded the best results for our model.

To measure the models performance, we analyzed both loss and accuracy. To measure loss, we utilized Cross Entropy loss, a commonly used loss function for classification.

Finally, because the fine-tuned ResNet-50 model is so powerful and we wanted to avoid overfitting, the model was trained for 5 epochs.

5 Model results

5.1 Baseline model

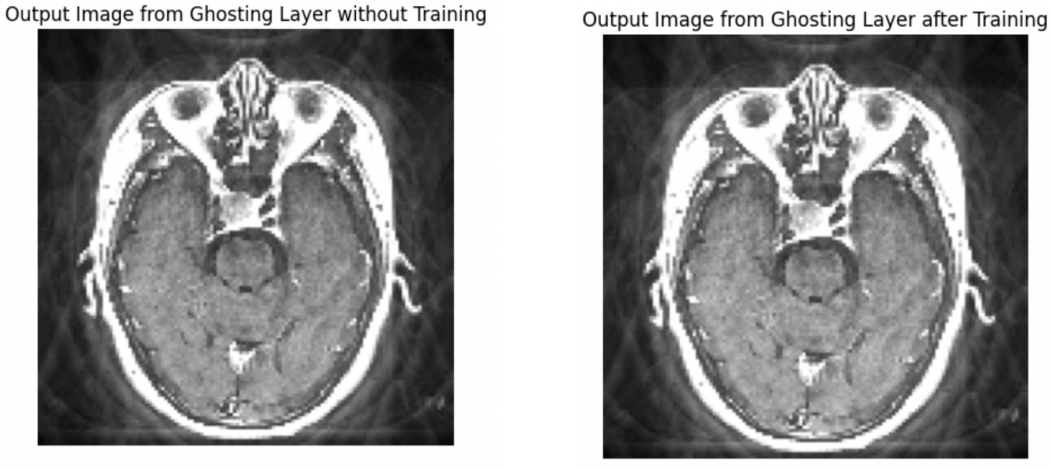
We first found a baseline accuracy, using the same model architecture, on our original images without any ghosting artifacts. This resulted in a 93.75% classification accuracy on the test set. Our target accuracy for our ghosted classification model was equivalent to or better performance than this original model.

5.2 Ghosted model weights

After training our model with the 3 ghosting directions for five epochs, the final model weights were [1.0544565, 1.039232 , 1.0648373, 1.0468446] for the original image (no ghost), vertical ghost,

horizontal ghost, and diagonal ghost, respectively. Therefore, all of the weights increased from their original weight of 1, making each of the ghosts more prominent in the final prediction from the physical layer after training. Below is an example of the ghosting layer output on an image from the training dataset, before and after training.

Figure 5: Ghosting Layer Output Comparison



Based on the model weights, the ranking of ghost directions from most to least important for classification is as follows: horizontal, no ghost, diagonal, and vertical ghost. Therefore, horizontal ghost had the most importance for classification accuracy while vertical ghost had the least importance.

5.3 Ghosted loss and accuracy

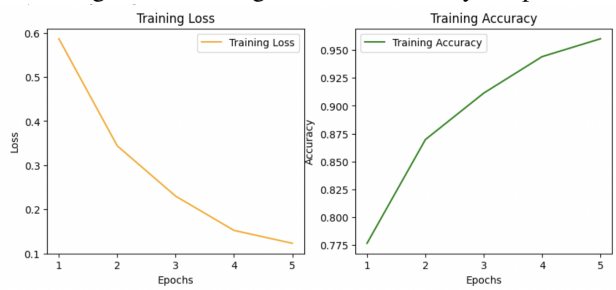
Below is a table of the training loss and accuracy throughout the five epochs.

Table 1: Training Loss and Accuracy

Epoch	Training Loss	Training Accuracy
1	0.5869	77.64%
2	0.3440	86.96%
3	0.2299	91.14%
4	0.1521	94.42%
5	0.1229	96.02%

The training loss was steadily decreasing while the training accuracy was steadily increasing. To better visualize this, we've provided a graph of the loss and accuracy for this model.

Figure 6: Training Loss and Accuracy Graphs

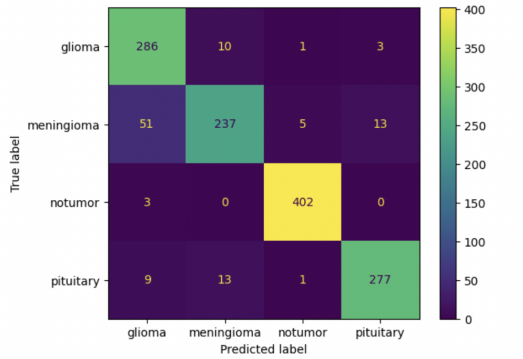


After training, the model was evaluated on the unseen test dataset. The final test loss and accuracy were **0.2389** and **91.69%**, respectively. With a comparable accuracy to the original dataset, this is demonstrating the viability of the ResNet-50 model on our unique dataset.

5.4 Ghosted confusion matrix

In real-world scenarios, it is necessary to analyze the specific performance on each of the four classes, as it is important to ensure that the model is misclassifying a tumor as a non-tumor with low probability. Below is a visual of the confusion matrix to further analyze the performance of the model.

Figure 7: Confusion Matrix Test Set



With a total of 1311 images and 906 of those as tumors, only 7 (0.77%) were misclassified as no tumor, proving the model's success of differentiating between tumor and no tumor. The model had the most difficulty with classifying meningioma tumors, as 69 images (22.54%) were classified incorrectly, with 51 of those being classified as glioma. As for the other classes, only 14 of 306 (4.67%) glioma-labeled samples were misclassified, 3 of the 405 (0.74%) no tumor samples were misclassified, and 23 of the 300 (7.67%) pituitary samples were misclassified.

6 Future work

Given more time, we'd like to explore other models with different architecture, such as ResNet18, ResNet101, and VGG16 (Visual Geometry Group). It would be interesting to assess how varying depths and model architecture affect the loss and accuracy performance. Finally, we prescribed the ghosting artifacts in this work and trained the neural network for the optimal weights for a weighted sum of ghosting artifacts for further classification, whereas ideally, the algorithm achieves optimal performance by also deciding the best motion artifacts to apply to the original image to amplify important, tumor-related features. One could, for example, explore the route of training motion artifact simulation parameters. In our case, we prescribed the modulation magnitude, kernel width, and periodicity, whereas in an ideal model, the algorithm would be able to train for these parameters individually for each direction of amplitude modulation- leading to a total of 12 trainable weights, 6 of which would have a non-negative physical constraint. We anticipate improved model performance with this new physical layer. On the other hand, it might also be worth it to include additional physical layers to add motion artifacts directly in the spatial domain.

7 Conclusion

Horizontal ghosting was the most important for our model's classification accuracy with the highest post-training model weight of 1.0648373. All ghosting direction weights increased from the initial value of 1, indicating that some level of ghosting was important to our model's classification accuracy rather than an image without ghosting at all. The testing accuracy of our model was 91.69%, which was only 2.06% off of our baseline model. This implicates high performance of our model for accommodating 3 different forms of noise for tumor classification. This could move future machine

learning applications on ghosting away from ghosting removal to ghosting accommodation in model training.

References

[American Cancer Society, 2024] American Cancer Society (2024). Survival Rates for Selected Adult Brain and Spinal Cord Tumors — cancer.org. https://www.cancer.org/cancer/types/brain-spinal-cord-tumors-adults/detection-diagnosis-staging/survival-rates.html. [Accessed 02-05-2024].

[Bhuvaji et al., 2020] Bhuvaji, S., Kadam, A., Bhumkar, P., Dedge, S., and Kanchan, S. (2020). Brain tumor classification (mri). Kaggle.

[Cheng, 2017] Cheng, J. (2017). Brain tumor dataset. figshare.

[Gallagher et al., 2008] Gallagher, T. A., Nemeth, A. J., and Hacein-Bey, L. (2008). An introduction to the fourier transform: Relationship to mri. *American Journal of Roentgenology*, 190(5):1396–1405. PMID: 18430861.

[Hamada, 2020] Hamada, A. (2020). Br35h: Brain tumor detection 2020. Kaggle.

[He et al., 2015] He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, abs/1512.03385.

[King et al., 1995] King, K. F., Crawford, C. R., and Maier, J. U. (1995). Correction for filter-induced ghosts in echo planar imaging. In *Proc Soc Magn Reson Med*, volume 3, page 105.

[Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.

[Kyathanahally et al., 2018] Kyathanahally, S. P., Döring, A., and Kreis, R. (2018). Deep learning approaches for detection and removal of ghosting artifacts in mr spectroscopy. *Magnetic Resonance in Medicine*, 80(3):851–863.

[Nickparvar, 2021] Nickparvar, M. (2021). Brain tumor mri dataset. Kaggle.

[Preston, 2016] Preston, D. C. (2016). MRI&xA0; Basics — case.edu. https://case.edu/med/neurology/NR/MRI%20Basics.htm. [Accessed 01-05-2024].

[Reeder et al., 1997] Reeder, S. B., Atalar, E., Bolster Jr, B. D., and McVeigh, E. R. (1997). Quantification and reduction of ghosting artifacts in interleaved echo-planar imaging. *Magnetic resonance in medicine*, 38(3):429–439.

[SEER, 2024] SEER (2024). Cancer of the Brain and Other Nervous System - Cancer Stat Facts — seer.cancer.gov.

[Zaitsev et al., 2015] Zaitsev, M., Maclaren, J., and Herbst, M. (2015). Motion artifacts in mri: A complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging*, 42(4):887–901.