Universiteit Leiden

Master Computer Science

Generating full-contrast T1ce MRI images from simulated low-dose T1 images using a neural field-based generative model to reduce Gadolinium dose.

Name:	Rianne Weber
Student ID:	s3703509
Date:	15/01/2024
Specialisation:	Bioinformatics
1st supervisor:	Lu Cao
2nd supervisor:	Marius Staring
3rd supervisor:	Yunjie Chen

Research report in Computer Science

Leiden University Medical Center, Division of Image Processing (LKEB)
Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

1 Abstract

To reduce the dosage of contrast agent necessary for accurate MRI imaging, a neural field-based model is trained on two datasets of MRI images of patients with brain tumors to generate full-contrast T1ce-weighted images from low-contrast T1. The low-contrast images are simulated using a linear transformation, and a linear transformation with added noise. Quantitative measures of the quality of the generated images are significantly affected by the noise added to the low-contrast images, but visually the two methods of low-contrast image generation are comparable. 25% dosage input leads to clearer images of the tumors than 10% dosage input, which is in line with expectations. Most predicted images are slightly darker or brighter than the true image, which should be investigated in future work. We conclude that neural field-based models have potential for use in clinical practice to generate full-contrast T1ce images from low-contrast images.

2 Introduction

Contrast agent is essential to MRI research. This agent is administered intravenously and subsequently changes the relaxivity of certain bodily structures [5], making these structures more visible in the images. However, there have been ongoing discussions about the health risks associated with Gadolinium-Based Contrast Agents (GBCA) [4], the most common group of contrast agents used for T1-weighted MRI imaging [5]. Deposition of Gadolinium in the brain may occur, which in rare cases may result in Nephrogenic Systemic Fibrosis [6]. The European Medicines Agency has recommended to cease the use of certain types of linear GBCAs altogether [1]. It is therefore important to investigate ways in which the dosage of this contrast fluid that is necessary for high-quality images may be reduced. One way of doing this is through deep learning methods, which have recently gained popularity in the field of radiology [12]. If the full-contrast images can be accurately generated, less GBCA is required in the imaging process. Several studies have focused on generating full-contrast MRI images from low-contrast ones [7, 8, 10, 15, 17]. In this study, we aim to generate full-contrast T1ce-weighted MRI images from low-contrast T1 images using a neural field-based network.

Mallio and colleagues [13] performed a literature review to investigate all articles on the topic of post-contrast simulation with Artificial Intelligence up until 2023. They conclude that while more research is necessary, the artificial generation of full-contrast MRI images has significant potential. An example of these studies is that of Pasumarthi and colleagues [17], who implemented a deep neural network to predict full-contrast images from a combination of pre-contrast T1 images and low-contrast ones. While in 10 percent of the cases, radiologists found different enhancing patterns in the reconstructed images, all of the diagnoses were the same as for the actual full-contrast image. This result is promising, but ideally, there would be even less discrepancy between the reconstructed images and the real full-contrast images. A similar model was used in two other studies [7, 8]. While the objective quality measures investigated by Haase and colleagues [8] yielded sufficient results, lesion areas were often not of high enough quality for the predicted images to be interchangeable with the real ones. Other networks, like a Bayesian deep learning model, have also been implemented on the same task [10]. The images generated with this model were quantitatively and qualitatively close to the real full-contrast images, but the authors mention that Generative Adversarial Networks may yield even better

results.

Chen and colleagues [3] proposed a version of a GAN named Conditional Neural fields with Shift modulation (CoNeS) and tested it on two datasets: BraTS2018 and a dataset of patients with a possible Vestibular Schwannoma. This study is different from previous research in that it works with a neural field-based model, which makes it more spacially aware than standard deep networks. Significant improvement compared to state-of-the-art models was found on multiple translation tasks, such as the reconstruction of T1ce from T1, T2, and FLAIR images. The model also outperformed others when predicting T1ce from T2 alone. However, the performance on the BraTS dataset was significantly better than on the Vestibular Schwannoma data, and a higher performance is necessary for clinical implementation. In this study, we extend this research by using the CoNeS model to predict T1ce images from low-contrast T1ce images. Using a low-contrast rather than a precontrast image may help improve the model's performance, since it contains more information about the areas that should be enhanced.

Some patterns can be found concerning the methodology of the existing research on this topic. Firstly, most previous research evaluates model performance using both overall quantitative measures, like the peak signal-to-noise ratio and the structural similarity index, and qualitative measures like the ability of radiologists to distinguish the original from the generated T1ce. In this study, we focus on quantitative measures of performance. Qualitative measures are out of the scope of the study. However, this will be added in future research since this may further substantiate clinical implementation.

Secondly, in most previous research the full-contrast image is predicted from not only the low-contrast one, but a combination of the low-contrast image with the pre-contrast T1, or the T2-weighted image. While generally more input data yields better results because the model has more information to work with, it is often the case that one or more modalities are not available [2]. Furthermore, the use of multiple modalities requires the registration of all modalities to each other, which takes time and may introduce inaccuracies. Therefore, the model ideally has as few input modalities as possible while retaining a high performance. To explore the quality of the prediction with as few modalities as possible, only the low-contrast T1 image will be used as input for our model.

Third, most research on the reduction of contrast agent uses real low-contrast images. However, taking these images is uncommon since it requires two separate injections and imaging sessions following each other directly. Therefore, to increase the amount of available training data, it is useful to develop a method to generate low-contrast images from the T1- and T1ce-weighted images. One other study was found where these images were generated [15]. This study approached the issue by lowering the signal-to-noise ratio (SNR) of the T1ce image. In our retrospective study, we combine this noise-based method with a linear transformation to obtain the low-contrast images. We also test a linear transformation on its own.

The purpose of this study is twofold: first, we apply both methods of generating low-contrast images to the datasets at hand and evaluate their differences. Then, we use the simulated low-contrast images to train the CoNeS model to reconstruct the full-contrast images and evaluate the results.

3 Methods

3.1 Model

The model used in this study is Conditional Neural fields with Shift modulation (CoNeS), as proposed by Chen and colleagues [3]. This is a Generative Adversarial Network (GAN) that uses Multi-Layer Perceptrons (MLP) rather than a Convolutional Neural Network (CNN), to prevent the development of spectral biases. Furthermore, this model uses neural fields, meaning the target images are represented as continuous coordinates rather than discrete values like in a standard CNN. The model is conditional because the neural fields are conditioned on latent information in the input images that is extracted with a hypernetwork. Shift modulation refers to the linear mapping of the parameters of all neurons in an MLP so that the parameters do not have to be generated pixel-wise. Shift modulation is computationally less costly and is less likely to lead to overfitting. In this study, the model was run for 250 epochs during training.

3.2 Datasets

Two datasets were used separately to train the model. The first is the train set of the BraTS 2020 dataset [14], containing images of patients with brain tumors in 4 different modalities (T1, T1ce, T2, FLAIR) as well as the segmentation of the tumor. BraTS 2020 provides a separate validation set, but this set was not used so that it can be used as a test set in future research.

Characteristics	Vestibular Schwannoma		BraTS 2020	
Section thickness	1.0 mm		1.0 mm	
In-plane resolution	0.35 x 0.35 mm		1.0 x 1.0 mm	
	Raw data	Used in study	Raw data	Used in study
No. of patients	102	71	494	357
No. of imaging instances	433	175	494	357

Table 1: Statistics of the datasets.

Secondly, a dataset consisting of patients with a possible Vestibular Schwannoma (VS), a benign tumor, was used [16]. Some patients from this dataset underwent multiple scanning sessions, to track the possible growth of the Vestibular Schwannoma. These scanning instances of the same patient were treated as separate data instances for the model since the date of the scan is not relevant in our case. However, separate imaging instances of the same individual were not split between the train and test sets (see 3.7). An overview of the datasets can be seen in Table 1. Figure 1 displays an overview of the preprocessing pipeline. All further steps of this pipeline will be discussed in the next paragraphs.

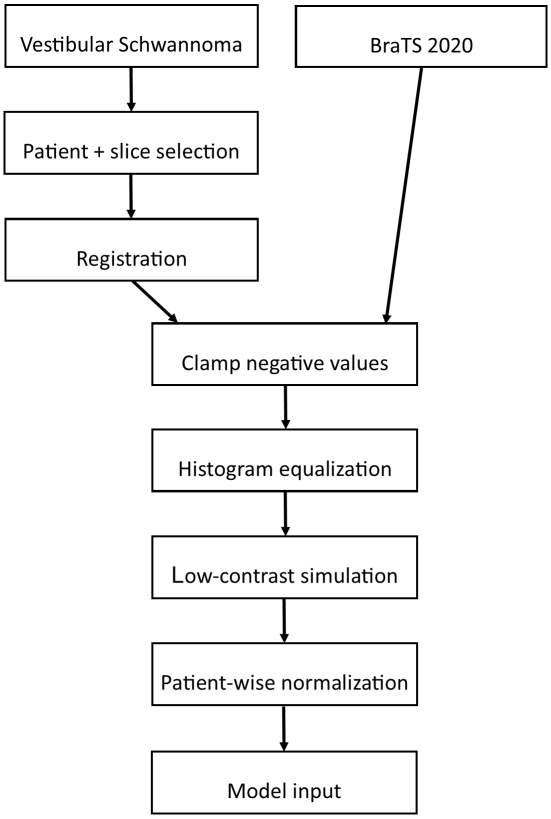


Figure 1: Preprocessing pipeline of the VS and BraTS 2020 datasets.

3.3 Registration and selection

The BraTS dataset is preregistered, but the VS dataset was not yet registered. Rigid registration was applied using the Python extension of Elastix [11]. The resulting registered images were all manually checked for inconsistencies, and slices with missing data were removed. 12 images were completely left out due to insufficient registration or image quality. Some VS data instances did not have both the T1 and T1ce-weighted images, in which case these were also removed.

3.4 Histogram equalization

Most of the pairs of T1 and T1ce images, especially for the BraTS dataset, had different intensity ranges. One extreme example is that of a T1 image with a maximum intensity of 25000, while its complementing T1ce image had a maximum of 800. This is undesirable, since it means the entire image is darker or lighter, not only the contrast-enhanced areas. Therefore, the model will learn to darken or lighten the image as a whole. To prevent this as much as possible, histogram equalization was applied. First, histogram matching was tested. With histogram matching, a mapping function is applied to the intensity histogram of one image to exactly match that of a second given image. While this method worked, it was considered too extreme since pixel values are individually altered. A less extreme method that was applied is scaling. The advantage of this method is that all pixels are scaled with the same value, meaning the intensity relationship between the pixels is maintained. The scaling was based on the peak of the intensity histograms of the T1 and T1ce images, which were determined using Scipy's *find_peaks* function. Different

hyperparameters of this function were tested to ensure at least one peak was found in each image and that the location of the peaks was as accurate as possible. The T1ce image was then scaled as follows:

$$T1ce_{scaled} = \frac{T1ce}{peak_{T1ce} * peak_{T1}} \tag{1}$$

Where $peak_{T1ce}$ and $peak_{T1}$ are the intensity values corresponding to the peaks determined previously. The assumption underlying this method is that the peaks in the intensity histograms of both the T1 and T1ce images refer to the non-enhanced areas of the images since the number of non-enhanced pixels is significantly larger than that of enhanced ones. From this, we then conclude that the pixels that form these peaks should be equal in intensity value. The enhanced areas are represented in the low-frequency, high-intensity values in the T1ce histogram. An example of the histogram equalization can be seen in Figure 2 and 3. The two rightmost images in Figure 2 are subtracted from respectively the original T1ce and the scaled T1ce and colorized to see the differences better. These two subtraction images should theoretically only display the enhanced areas of the image. The scaled subtraction image is therefore of higher quality than the original subtraction image. In Figure 3 the associated histogram is displayed. As is visible, the scaled T1ce image is more similar to the T1 than the original T1ce image.

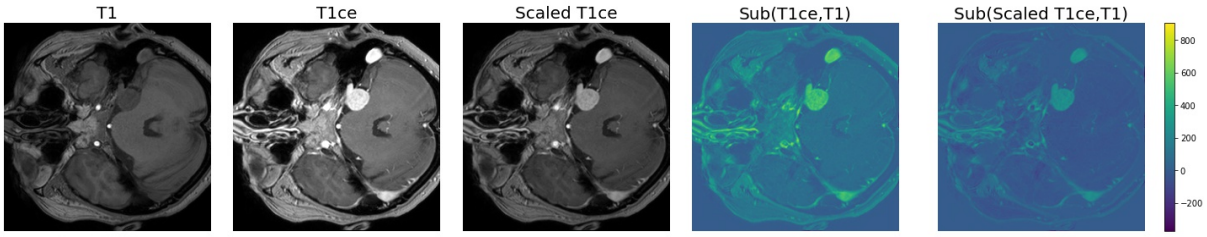


Figure 2: Example of histogram scaling. From left to right: Original T1, original T1ce, scaled T1ce, T1 subtracted from original T1ce, T1 subtracted from scaled T1ce. The last two images are colorized to see the differences better.

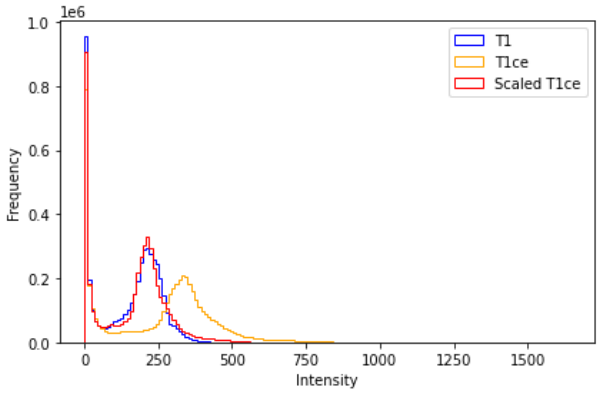


Figure 3: Histogram corresponding to the images from Figure 2. The scaled T1ce image is more similar to the T1 than the original T1ce image.

In some cases, the histogram equalization yielded worse results than the original images, which was either because the original images' histograms were already sufficiently close to each other, or because the wrong peaks were found. We manually checked all images after histogram equalization for quality control. If the equalized images were not of sufficient quality, it was checked whether the originals were better and if so, these were used instead. If neither the equalized nor the original images were of good quality, the image was left out altogether. For the Vestibular Schwannoma dataset, one image was left out. 12 images were removed from the BraTS dataset.

3.5 Generation of low-contrast images

In this study, we test two methods to generate low-contrast images: a simple linear transformation, and a linear transformation combined with noise based on the method found in previous literature [15]. We simulated 10% and 25% dosage.

The first method for generating the low-contrast image is through linear transformation. If an image with x% contrast is required, it would be obtained through the following formula:

$$x\%\text{contrast} = (1 - x) * T_1 + x * T_{1,ce} \tag{2}$$

This method is easy to apply to any dataset given that the T1 and T1ce-weighted images are available. It also does not require thorough preprocessing or meta-knowledge about the images that may be unavailable or costly to retrieve. Furthermore, it directly generates an image that is interpretable from a clinical standpoint. On the other hand, this method may be crude given that it assumes a linear relationship between dose and contrast. For the GBCA Gadopiclenol a linear relationship between dose and contrast-to-noise ratio was discovered between the concentrations of 0.025 mmol/kg and 0.10 mmol/kg [18]. Given that the standard administration of GBCA is around 0.10 mmol/kg [9], low-contrast images are of lower concentration than this, which would substantiate a linear approach to the simulation of low-contrast images. However, it should be noted that the linear relationship was determined based on only a few data points. We therefore do not assume the linear relationship as a proven truth, but rather as a reasonable possibility.

The second method of generating low-contrast MRI images has been found in previous literature [15]. In this study, the subtraction image (T1ce - T1) was lowered in quality intentionally to simulate a lower dosage of contrast agent. This lower-contrast subtraction image was then used as input for the model. More specifically, given lower contrast percentage f and low-contrast subtraction image sub_{lc}, Gaussian noise was added to the subtraction image as follows:

$$sub_{lc} = sub + N(\sigma^2, 0) \tag{3}$$

The standard deviation of this Gaussian noise σ is calculated as

$$\sigma = \sigma_{T_1} * \sqrt{2} * \sqrt{\frac{1}{f^2} - 1} \tag{4}$$

In this formula, σ_{T_1} refers to the standard deviation of the noise of the T1-weighted image. Since this method assumes that the noise levels in the T1 and T1ce images are identical, the noise of T1ce can also be used instead.

As is visible from the formula, this method depends on the Signal-to-Noise Ratio (SNR) of the original images, which may be difficult to estimate. We used an SNR of 50 as suggested by the authors of the article. Another notable feature of the method is that it works with the subtraction image rather than the 'real' low-contrast representation. The subtraction images cannot be used in real-life medical settings since they convey too little information about the surrounding bodily structures. The subtraction image could be added to the T1 image again, but this would not result in low-contrast images that are similar to real-life ones. We therefore added the noise to the image as a whole. By combining the linear method with this noise method, we aimed to create low-contrast images that are realistic both in terms of the brightness of the enhanced areas, as well as the noise of the overall image.

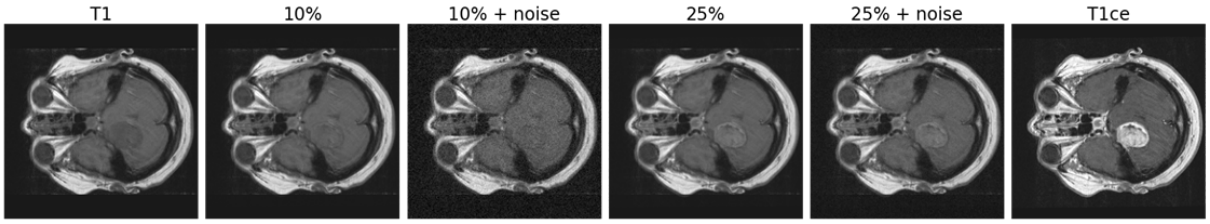


Figure 4: Different simulated low-contrast images for an image from the VS dataset. The images without noise were generated with the simple linear transformation. The ones with noise were made using a combination of the linear transformation and SNR reduction.

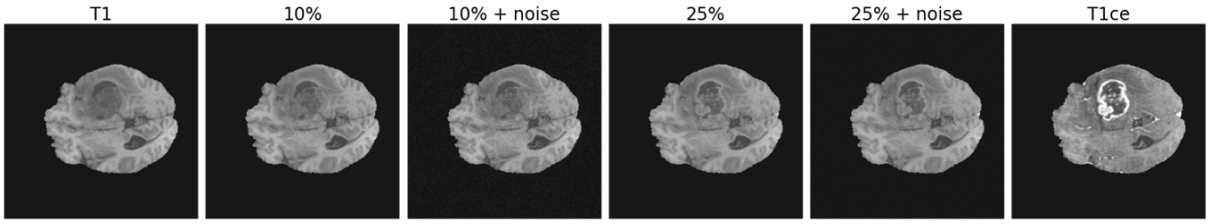


Figure 5: Different simulated low-contrast images for an image from the BraTS dataset.

Figure 4 and 5 display an example of the simulated low-contrast images for the VS and BraTS datasets, respectively. Visually, there does not seem to be a large difference between the simple linear images and the linear images with noise. We also calculated the Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM) of all input images compared to the original T1ce. The MSE is shown in Figure 6. The SSIM and PSNR display similar results, so these are not shown here, but they can be found in Appendix B. For both low-contrast simulation methods, the 25% dose image is closer to the T1ce than the 10% dose image. The methods with noise have a higher average MSE and a larger variance than those without. This is likely because measures like the MSE are sensitive to noise.

Despite the effect of the noise on the quantitative measures, the linear method with noise seems the most realistic in practice, since it takes into account both the change in brightness of enhanced areas as well as the noise of low-contrast images. However, the simple linear transformation yields images with quantitative measures that are more in line with the original T1 and T1ce. Because it would be interesting to see how the model's

performance changes based on whether there is noise in the input or not, we tested the model with both methods.

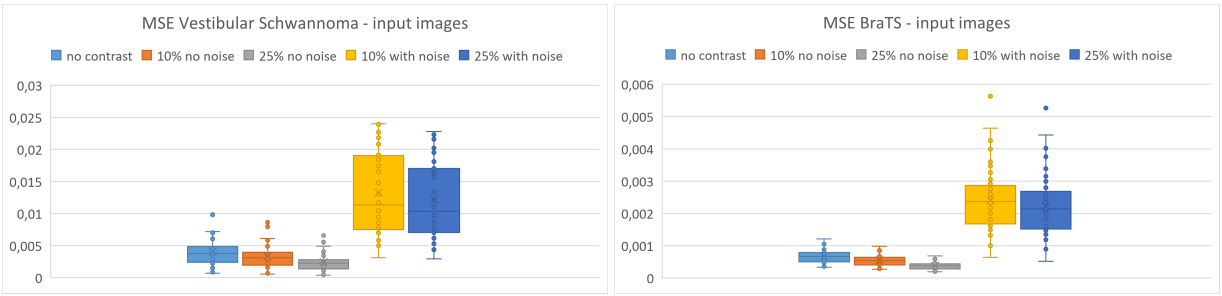


Figure 6: MSE of the real T1ce and the low-contrast input images for BraTS (right) and VS (left). *No contrast* refers to the original T1 image.

3.6 Normalization

Before using as input to the model, the image intensities were normalized between [-1,1]. At first, this was done for all images separately. However, due to the differing intensity ranges of the T1 and T1ce images from the same patient, this method of normalization scaled the images at different strengths and therefore shifted the peaks of the intensity histograms away from each other. In practice, this means that part of the histogram equalization would be undone. To counter this issue, patient-wise normalization was applied. Rather than using the maximum and minimum intensity values of each image separately to apply normalization, the maximum and minimum values were determined over the T1, T1ce and low-contrast images together. These values were then used to normalize all images. This method ensures the retention of the relative intensity frequencies. An example of this normalization method is displayed in Figure 7. As is visible, the non-enhanced areas of the patient-wise normalized images are significantly closer in brightness than the individually normalized images.

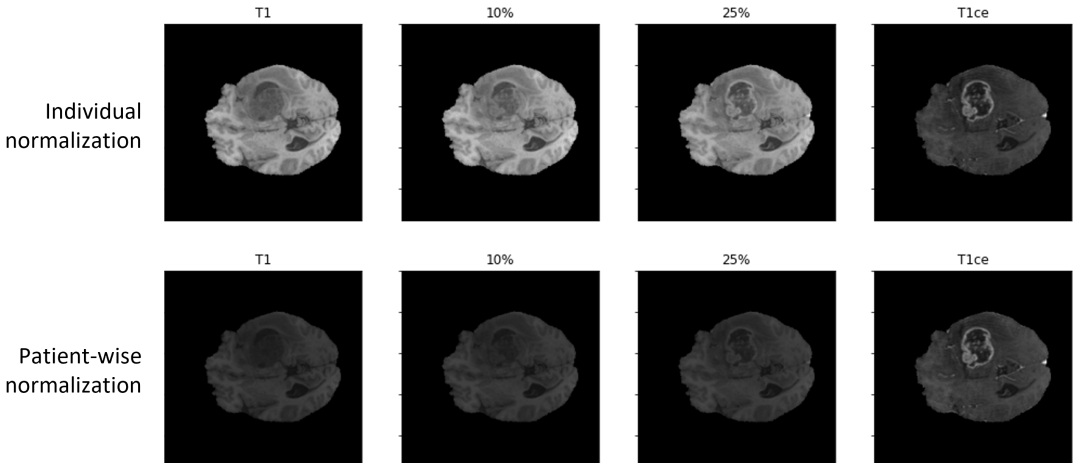


Figure 7: Example of individual normalization versus patient-wise normalization shown on the original T1, original T1ce, and linear low-dose simulations.

3.7 Training

Both datasets were first split into a train and test set. The BraTS data is presplit, and the VS set was manually split 90%/10%. The number of 2D slices per imaging session differed per patient, so splitting the data solely on the patient number would not be representative of an actual 90%/10% split. Therefore, the train-test split was based on the number of 2D slices rather than the patient number. Slices from the same patient were kept in the same set, to prevent introducing bias. This method resulted in 6041 slices in the train set and 624 slices in the test set. Both train sets were further split 80%/20% for validation. The test sets were not used in this study, so that they could be used in future research.

4 Results

First, the quantitative evaluation of the predicted T1ce images was performed. Table 2 displays the Mean Squared Error (MSE), Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) of the predicted images based on different datasets and contrast agent dosages as input. Boxplots of the SSIM are shown in Figure 8. The SSIM of the VS dataset follows the same pattern as when this measure is calculated for the input image itself. The 25% dose input performs better than the 10% dose for both the noise and no noise conditions. Furthermore, the model yields worse results when trained on noisy inputs compared to inputs without added noise. These same patterns can also be found in the PSNR and MSE, which is why these are not shown here (these can be found in Appendix A). However, for the BraTS dataset, the predicted T1ce from 10% dose without noise seems to deviate from this pattern. This may be due to random chance, but it warrants further investigation in the future. On average, the BraTS dataset yields better results than the VS set.

Data	MSE	SSIM	PSNR
BraTS 0%	0,00549±0,00221	0,91686±0,01631	23,03738±2,15439
BraTS 10%	0,00420±0,00148	0,74112±0,02134	24,06235±1,68868
BraTS 10% + noise	0,00842±0,00252	0,14263±0,02108	20,97765±1,51920
BraTS 25%	**0,00311±0,00115**	**0,94735±0,01001**	**25,48128±2,13011**
BraTS 25% + noise	0,00741±0,00153	0,15437±0,01958	21,40346±0,98867
VS 0%	0,00350±0,00280	0,76354±0,07294	25,52571±2,86196
VS 10%	0,00301±0,00272	0,79581±0,06531	26,31837±2,98394
VS 10% + noise	0,01569±0,00928	0,51210±0,10167	18,63893±2,23737
VS 25%	**0,00232±0,00241**	**0,83471±0,05472**	**27,55205±2,98896**
VS 25% + noise	0,01379±0,00686	0,56461±0,11174	19,06021±2,02142

Table 2: Result of quantitative evaluation of model performance given the two datasets and different contrast dosages as input. The values in the table represent mean ± standard deviation. The best value in each column per dataset has been highlighted in **bold**.

Figure 9 displays an example of the input, predicted, and real full-contrast images for VS and BraTS, respectively. Two more examples can be found in Appendix C. Notably, the BraTS predicted images are significantly brighter than their corresponding T1 and T1ce. This is less visible for the VS data. It seems the intensity values of the non-enhanced

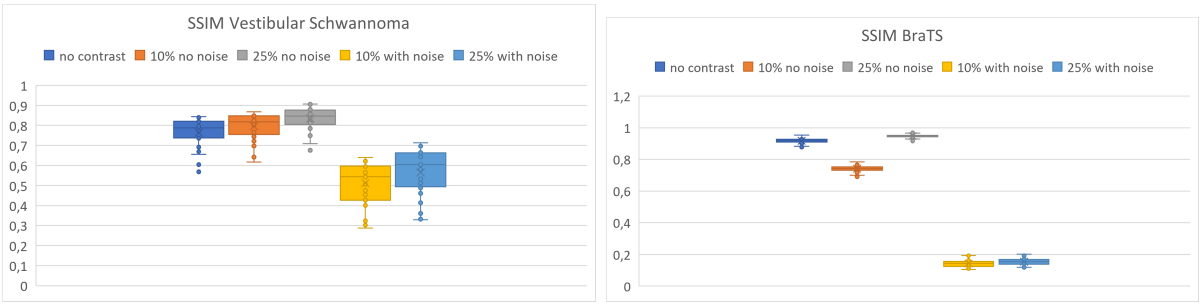
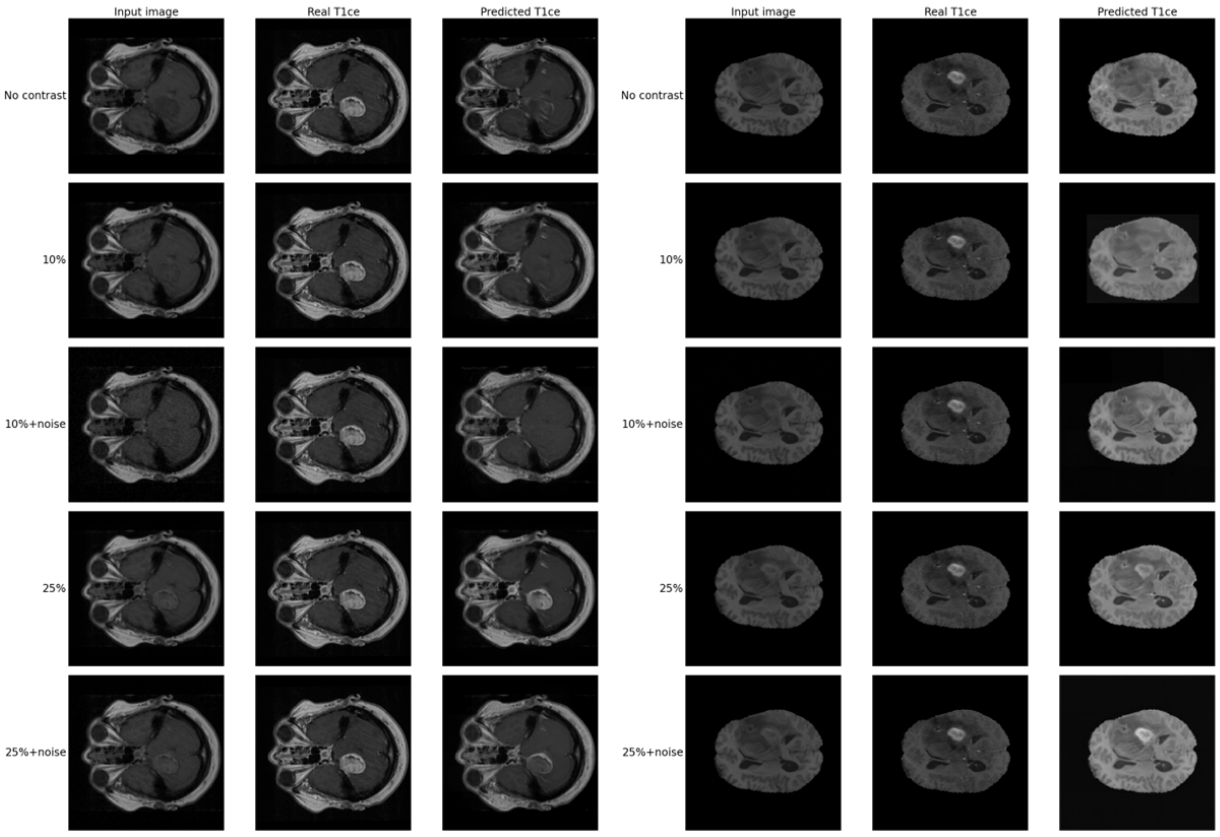


Figure 8: SSIM of the real T1ce and the predicted T1ce for BraTS (right) and VS (left) based on different types of low-contrast inputs.

areas are overestimated, which may be due to the low input intensities or suboptimal hyperparameters. For the 25% and 25%+noise inputs, this difference is less pronounced. The tumors are usually most clear in the images generated with 25% dosage as input. In some cases, like the example in Figure 9a, the tumor is not detected with 10% dosage input. The models with 25% and 25%+noise input are visually more similar to each other than the MSE, PSNR and SSIM corresponding to the same images would suggest. It is possible that a very small amount of noise remains that is too weak to see, but still affects the quantitative results.



(a) VS (b) BraTS

Figure 9: Example of input, prediction and real images for both datasets.

5 Conclusion and Discussion

In this study, we first simulated low-contrast T1 image from a combination of T1 and T1ce images. The low-contrast images were simulated using two different methods: a linear transformation with noise and the same transformation without noise. The former method seems more realistic, given that it takes into account the noise that comes with administering only a small amount of contrast agent. However, since both have their advantages and drawbacks, both types of low-contrast images were used for the second part of the study.

Next, we used a neural field-based model to predict full-contrast T1ce images from the simulated low-contrast ones. Visually, the inputs without noise yielded slightly better results than the inputs with noise. The 25% inputs performed better than the 10% input across all conditions. The results for the BraTS dataset were better than for the VS dataset, which is in line with previous findings [3]. Given that in some cases the tumor was not detected with 10% dosage input, we conclude that a 25% dosage input is more viable for clinical implementation than 10%.

This study has a few limitations that should be kept in mind. First and foremost, since this is a retrospective study we did not have access to real low-contrast images and therefore generated these ourselves. These images may slightly differ from the real ones, meaning a generated image with 10% contrast may not be the same as a real image with 10% contrast. This problem was addressed by testing different methods for generating these images to find the method that most closely resembles real-life low-contrast imaging. It should be noted that even when using real low-contrast images, it is difficult to determine the exact dosage that the image represents due to the variability of the exact time between contrast agent injection and imaging.

Second, the evaluation of the full-contrast images generated by the model was mainly done using quantitative measures of the entire image and visual confirmation. More measures like evaluation by radiologists or comparisons of true and predicted tumor segmentations were not feasible to include in the current study, but we aim to add this in the future.

Third, only the 10% and 25% contrast images have been investigated in the current study. We have made a start with determining the relationship between dosage and quality of the generated image, but it would be insightful to further clarify this relation by using more dosages as input for the model.

We further aim to improve the model by tuning its hyperparameters and improving the normalization method of the images. This way, the overall brightness of the predicted images will likely not differ as much from the true T1ce images as it did in the current study. Lastly, we plan on applying k-fold cross-validation to the datasets to increase the accuracy of the results in the future.

Reflection

In the last couple of months I have been able to experience working as a researcher for the Division of Image Processing (LKEB) at the Leiden University Medical Center (LUMC). I enjoyed being part of a group of people who are all focused on improving medical procedures for patients. The work in this field has direct practical relevance, which is very motivating.

During my internship, I learned a lot. First, I have become more familiar with the methods used to process MRI images, as well as the process of managing MRI data. Besides this, I also have learned more about my functioning in a job as a researcher. I enjoy this work and I am considering a career in bioinformatical science, starting with a PhD.

I quickly noticed that the process of doing research is less linear than I had initially expected. Some days and weeks, things ran smoothly and I felt I was making big steps, while at other times I felt I needed a lot of time for a seemingly easy task. My planning therefore had to be changed a lot. This issue was solved with some flexibility, and by planning a little more time for tasks than they were expected to take. In line with this, I have learned that the preprocessing of the data is an essential part of doing research and that this process can take a lot of time. At the start of the project, I wanted to be done with the preprocessing part as soon as possible because I felt that the 'real work' started once this was finished. However, the preprocessing is an important step of the research, and I kept returning to it even after doing model runs. In a future research project, I will not expect to have preprocessed the data instantly, and I will be aware that preprocessing problems can occur even out of my control.

In conclusion, I have enjoyed this internship very much and I have realized that this kind of job might be something for me in the future. I look forward to continuing this project for my thesis.

References

[1] European Medicines Agency. Ema's final opinion confirms restrictions on use of linear gadolinium agents in body scans. `https://www.ema.europa.eu/en/news/emas-final-opinion-confirms-restrictions-use-linear-gadolinium-agents-body-scans`, July 2017. Accessed: 01-12-2023.

[2] Reza Azad, Nika Khosravi, Mohammad Dehghanmanshadi, Julien Cohen-Adad, and Dorit Merhof. Medical image segmentation on mri images with missing modalities: A review, 2022.

[3] Yunjie Chen, Marius Staring, Jelmer M. Wolterink, and Qian Tao. Local implicit neural representations for multi-sequence mri translation. *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, 2023.

[4] Jin Woo Choi and Won-Jin Moon. Gadolinium deposition in the brain: Current updates. *Korean Journal of Radiology*, 20(1):134, 2019.

[5] Luis M. De León-Rodríguez, André F. Martins, Marco C. Pinho, Neil M. Rofsky, and A. Dean Sherry. Basic mr relaxation mechanisms and contrast agent design. *Journal of Magnetic Resonance Imaging*, 42(3):545–565, 2015.

[6] R. Garrick. Nephrogenic systemic fibrosis: Suspected causative role of gadodiamide used for contrast-enhanced magnetic resonance imaging. *Yearbook of Medicine*, 2007:217–218, 2007.

[7] Enhao Gong, John M. Pauly, Max Wintermark, and Greg Zaharchuk. Deep learning enables reduced gadolinium dose for contrast-enhanced brain mri. *Journal of Magnetic Resonance Imaging*, 48(2):330–340, 2018.

[8] Robert Haase, Thomas Pinetz, Zeynep Bendella, Erich Kobler, Daniel Paech, Wolfgang Block, Alexander Effland, Alexander Radbruch, and Katerina Deike-Hofmann. Reduction of gadolinium-based contrast agents in mri using convolutional neural networks and different input protocols. *Investigative Radiology*, 58(6):420–430, 2023.

[9] Micheal A. Ibrahim, Bita Hazhirkarzar, and Arthur B. Dublin. Gadolinium magnetic resonance imaging, Jul 2023.

[10] Jens Kleesiek, Jan Nikolas Morshuis, Fabian Isensee, Katerina Deike-Hofmann, Daniel Paech, Philipp Kickingereder, Ullrich Köthe, Carsten Rother, Michael Forsting, Wolfgang Wick, and et al. Can virtual contrast enhancement in brain mri replace gadolinium? *Investigative Radiology*, 54(10):653–660, 2019.

[11] Stefan Klein, Marius Staring, Keelin Murphy, Max A. Viergever, and Josien P. W. Pluim. elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, 2010.

[12] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[13] Carlo A. Mallio, Alexander Radbruch, Katerina Deike-Hofmann, Aart J. van der Molen, Ilona A. Dekkers, Greg Zaharchuk, Paul M. Parizel, Bruno Beomonte Zobel, and Carlo C. Quattrocchi. Artificial intelligence to reduce or eliminate the need for gadolinium-based contrast agents in brain and cardiac mri. *Investigative Radiology*, 58(10):746–753, 2023.

[14] Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, and et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, Dec 2014.

[15] Gustav Müller-Franzes, Luisa Huck, Soroosh Tayebi Arasteh, Firas Khader, Tianyu Han, Volkmar Schulz, Ebba Dethlefsen, Jakob Nikolas Kather, Sven Nebelung, Teresa Nolte, and et al. Using machine learning to reduce the need for contrast agents in breast mri through synthetic images. *Radiology*, 307(3), 2023.

[16] Olaf M. Neve, Yunjie Chen, Qian Tao, Stephan R. Romeijn, Nick P. de Boer, Willem Grootjans, Mark C. Kruit, Boudewijn P. Lelieveldt, Jeroen C. Jansen, Erik F. Hensen, and et al. Fully automated 3d vestibular schwannoma segmentation with and without gadolinium-based contrast material: A multicenter, multivendor study. *Radiology: Artificial Intelligence*, 4(4), 2022.

[17] Srivathsa Pasumarthi, Jonathan I. Tamir, Soren Christensen, Greg Zaharchuk, Tao Zhang, and Enhao Gong. A generic deep learning model for reduced gadolinium dose in contrast-enhanced brain mri. *Magnetic Resonance in Medicine*, 86(3):1687–1700, 2021.

[18] Philippe Robert, Véronique Vives, Anne-Laure Grindel, Stéphane Kremer, Guillaume Bierry, Gaelle Louin, Sébastien Ballet, and Claire Corot. Contrast-to-dose relationship of gadopiclenol, an mri macrocyclic gadolinium-based contrast agent, compared with gadoterate, gadobenate, and gadobutrol in a rat brain tumor model. *Radiology*, 294(1):117–126, 2020.

Appendix A

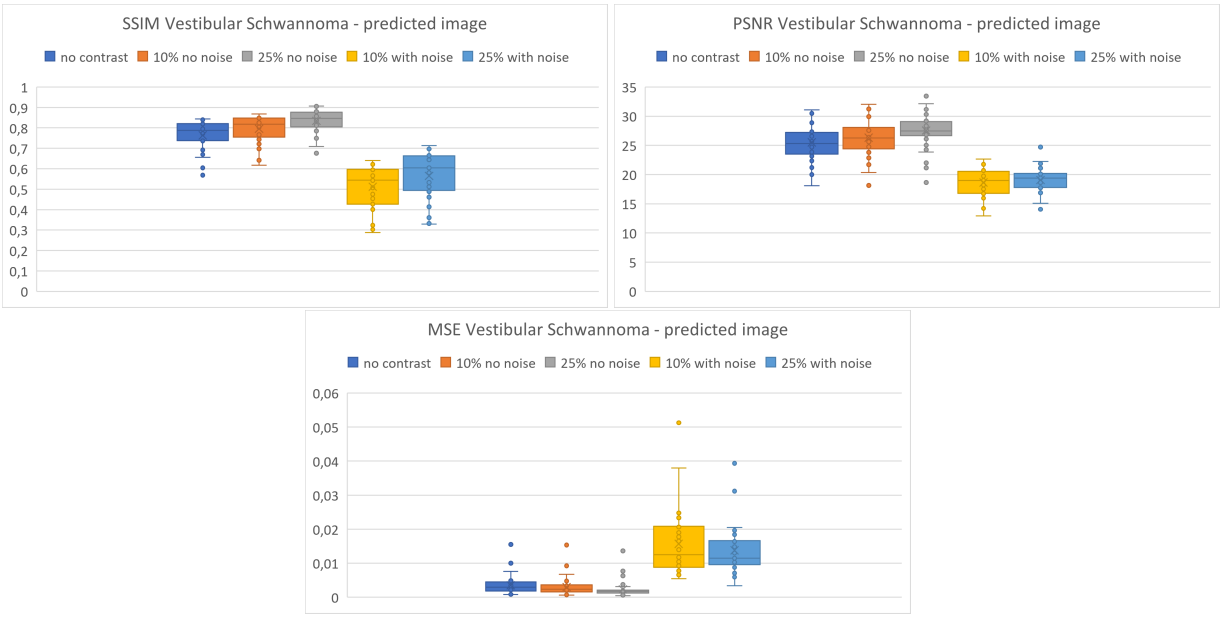


Figure 10: Quantitative evaluation of model using the VS dataset as input. Each boxplot refers to the SSIM, PSNR, or MSE measured on the actual T1ce and the predicted T1ce based on different types of inputs.

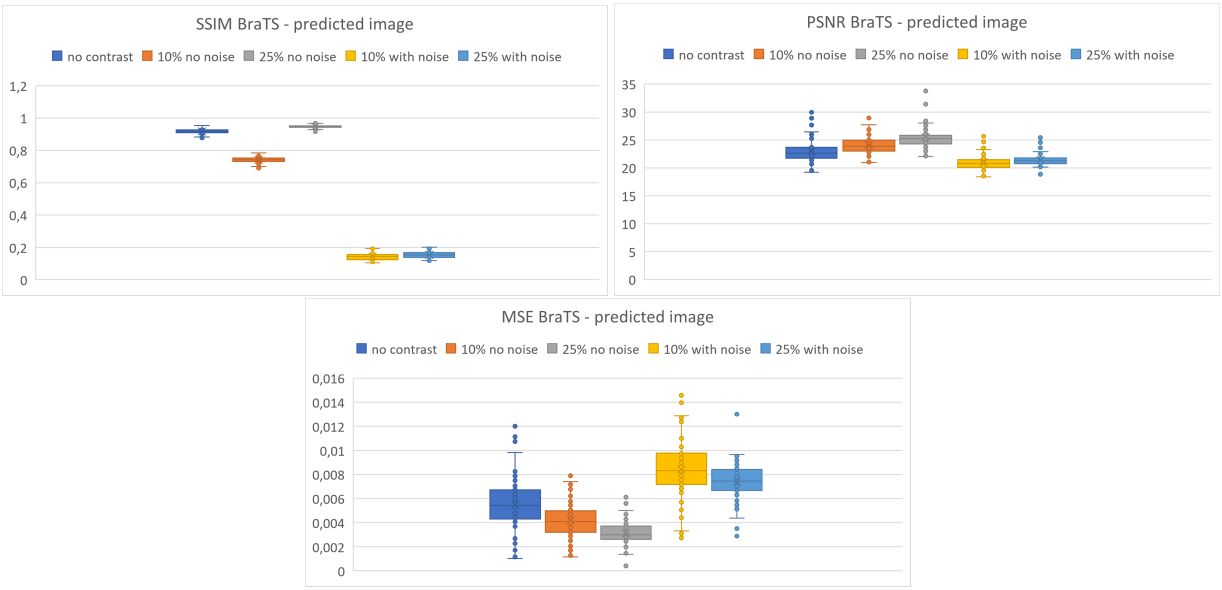


Figure 11: Quantitative evaluation of model using the BraTS dataset as input. Each boxplot refers to the SSIM, PSNR, or MSE measured on the real T1ce and the predicted T1ce based on different types of inputs.

Appendix B

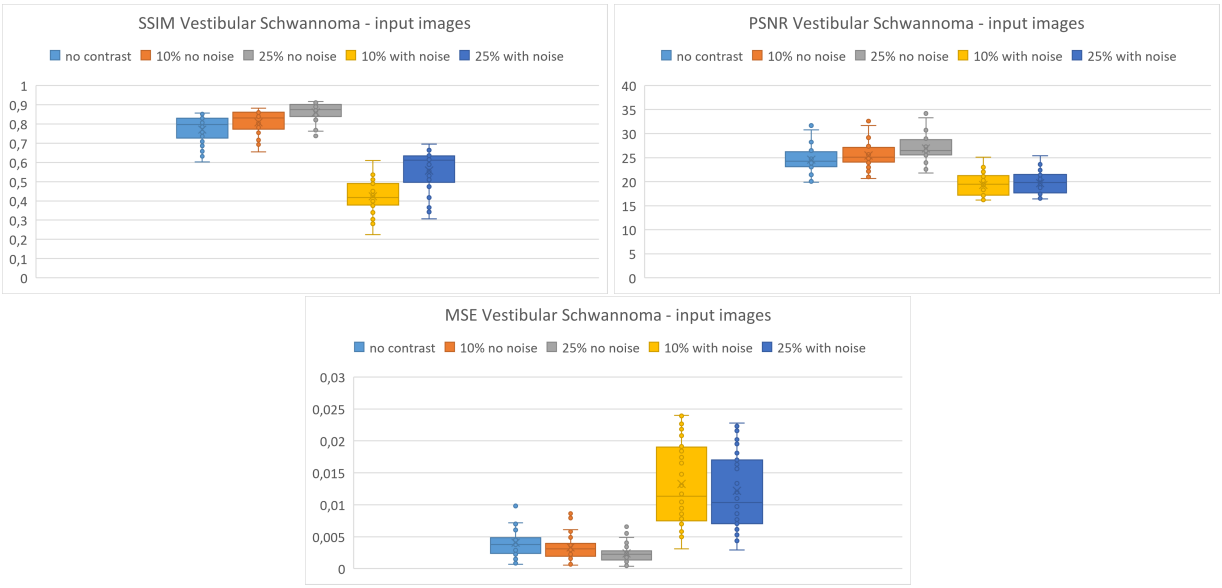


Figure 12: Quantitative evaluation of input images from the VS dataset. Each boxplot refers to the SSIM, PSNR, or MSE measured on the real T1ce and the different types of input images.

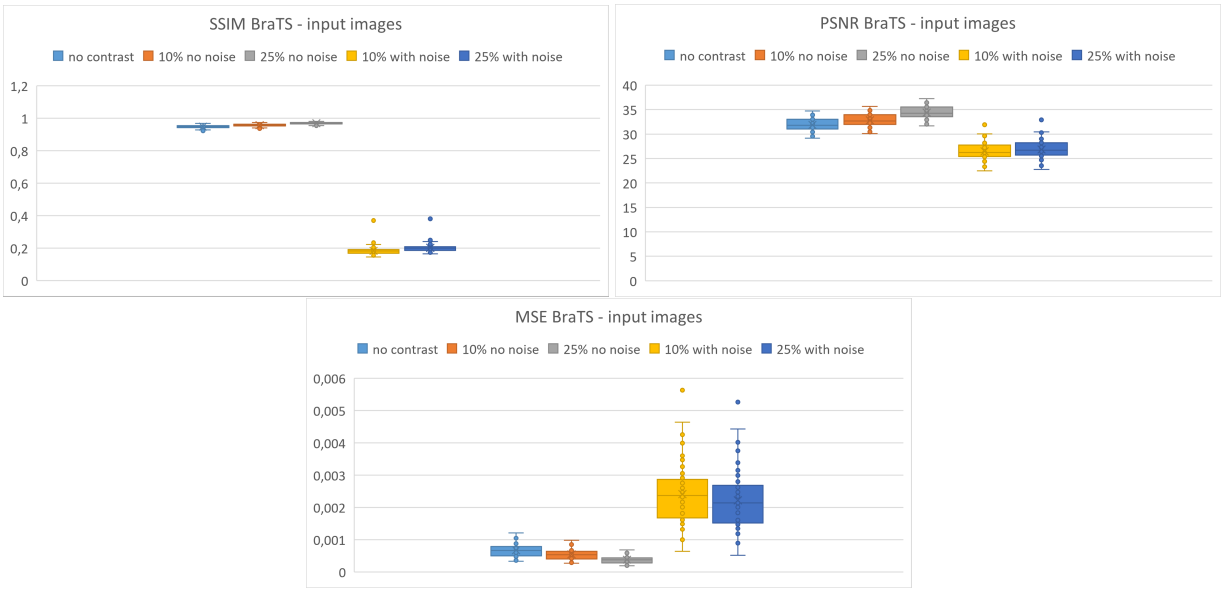
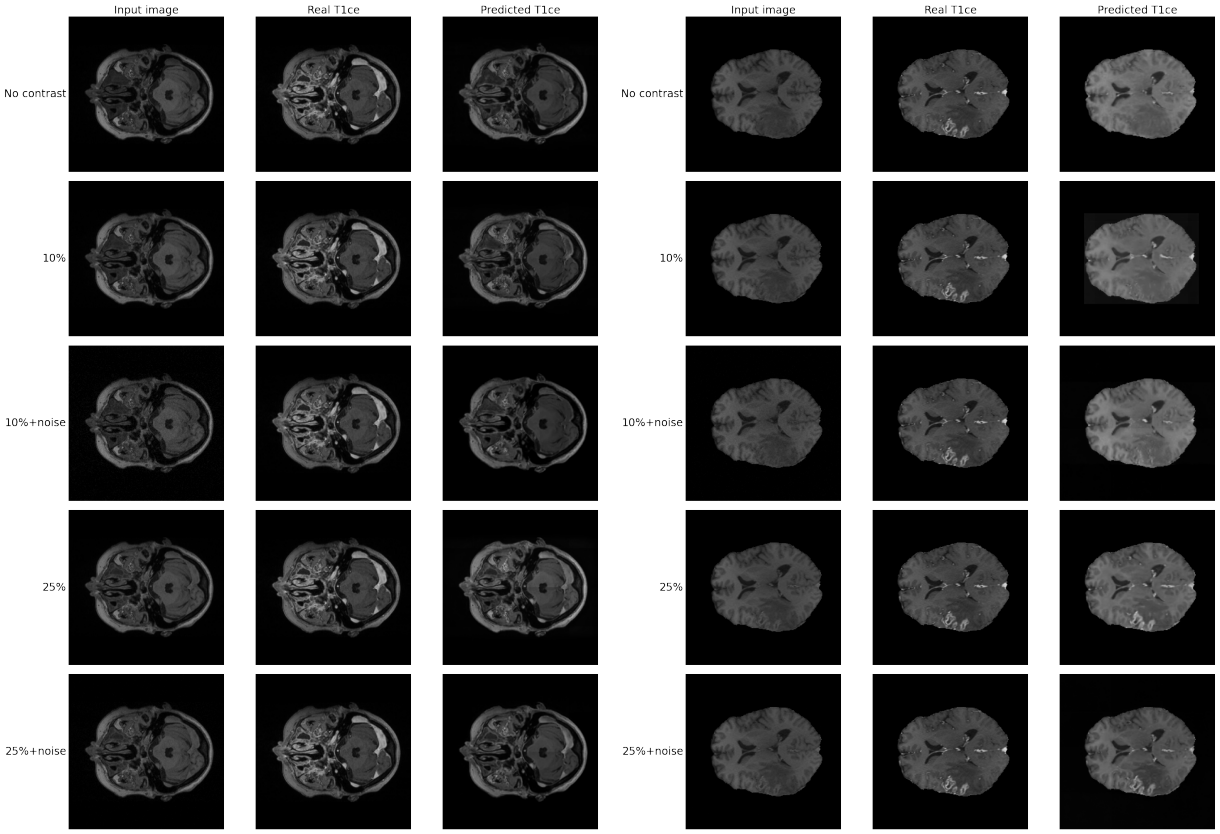


Figure 13: Quantitative evaluation of input images from the BraTS dataset. Each boxplot refers to the SSIM, PSNR, or MSE measured on the real T1ce and the different types of input images.

Appendix C



(a) VS (b) BraTS

Figure 14: Further example of input, prediction and real images for both datasets.