



TECHNICAL UNIVERSITY OF MUNICH

MLMI: MACHINE LEARNING IN MEDICAL IMAGING STARTSEITE

Final Report

Inpainting in Medical Imaging

Authors:

Alp Güvenir
Egemen Kopuz
Fatbardh Feta
Mahaut Gerard

Tutors:

Azade Farshad
Yousef Yeganeh

Winter 2021-2022

CONTENTS

Contents	1
1Introduction	2
2Background	2
2.1 Tumor Labels and Multi-modal MR Images	2
2.2 Tumor Inpainting	2
3Related Work	2
3.1 Conditional Anomaly Generation	2
3.2 Objective Functions	3
4Methodology	3
4.1 Concentric Approach to Mask Generation	3
4.1.1 Computing a concentric circle	3
4.1.2 Value Noise	3
4.1.3 Radius Noise	3
4.1.4 Ablation Components	3
4.2 Blob Approach to Mask Generation	3
4.2.1 Blob shape reconstruction	3
4.2.2 Tumor level segmentation	4
4.3 Tumor Inpainting	4
4.4 Tumor Inpainting using alternative architectures	5
4.5 Brain tissue deformation	5
5Experiments	5
5.1 Dataset	5
5.2 Evaluation Metrics	5
5.3 Concentric Approach	6
5.3.1 Binary Tumor Mask	6
5.3.2 Tumor Label Mask	6
5.4 Blob Approach	6
5.4.1 Blob shapre reconstruction	6
5.4.2 Tumor level segmentation	6
5.5 Tumor Inpainting	6
5.5.1 Focal Frequency Loss	6
5.5.2 Pix2Pix Architecture	8
6Discussion	10
6.1 MRI synthesis	10
7Conclusions	11
8Future Work	11
Acknowledgments	13
References	13
A.1 A concentric circle with sampled radiuses	14
B.2 Visual Validation Results for Binary Tumor Mask Generation	15
B.3 Validation Confusion Matrix for Binary Tumor Mask Generation	16
C.4 Visual Validation Results for Tumor Label Mask Generation	17
C.5 Validation Confusion Matrix for Tumor Label Mask Generation	18

Inpainting in Medical Imaging

GÜVENIR, KOPUZ, FETA AND GERARD



Fig. 1. Our work can be divided into two parts, namely: mask generation and inpainting. In the mask generation part, we try to produce binary masks aiming to form pathologically realistic looking tumour shapes. After that, we try to convert this binary mask to multi-class mask with tumour label information embedded. The main goal of these mask generation methods to achieve simplification while preserving tumour information as much as possible. In the inpainting part, we experiment with different network architectures and objective functions to improve upon chosen baselines [Isola et al. 2016; Kim et al. 2020].

1 INTRODUCTION

Anomaly detection and segmentation methods in medical imaging are often based on deep learning models which require a huge amount of balanced training data. However, medical image datasets are mostly imbalanced in terms of healthy and unhealthy samples. There is also a lack of pairs of healthy - unhealthy data. The aim of our project is to synthesize images of anomaly data, tackling both the imbalancedness of data, and the lack of healthy - unhealthy data pairs.

Image anomaly is defined as non-conforming patterns with respect to the image structure. Therefore, anomalies don't stand within the standard distribution of the data. In the medical domain, anomalies are common, and this is often what we want to detect when doing disease detection. There are various types of anomalies, in various kind of organs, and there are several types of medical imagery techniques. In our project, we focused only on MRI upper slice of brain tumor. It is also important to keep in mind that we want to have realistic medical MRI at the end, with realistic looking tumors.

2 BACKGROUND

2.1 Tumor Labels and Multi-modal MR Images

During this project we have used multi-modal MRI scans of brains from different individuals who have developed brain tumor. Multi-modal MRI scans consist of 4 different channels, namely T1, T1ce, T2 and Flair. Those channels are used to better represents different bodies and qualities present in the brain for example: presence of

gray matter, liquids , tumors, etc. During the length of this project we have worked with horizontal 2D cuts of those images

In our datasets tumours are segmented and divided into three layers. The outermost layer is the edema layer then there is the enhancing tumour layer. The innermost layer is called the necrotic layer and is the nucleus of the tumour.

2.2 Tumor Inpainting

Inpainting techniques simply add new elements into the original image in such a way that the inpainted region cannot be detected by a causal observer. It is a image editing technique that Our goal is to use inpainting techniques to insert tumours into the MRI images of brains.

3 RELATED WORK

3.1 Conditional Anomaly Generation

[Kim et al. 2020] proposes a pipeline that synthesises tumours in multi-contrast brain MR images. You may know that tumours in the brain can have really complex features depending on many factors; therefore, the authors came up with a simplification method which reduces this brain tumour information to concentric circles. With such concentric circles, one can now generate corresponding masks. At the end, they perform inpainting using the generated masks in fusion-based generative network. We have used this pipeline together with its philosophy as our foundation in our many experiments you will later see in the presentation.

[Isola et al. 2016] proposes an architecture benefiting from conditional adversarial networks with the purpose of using for image-to-image translation problems. It was demonstrated that this approach is effective at synthesizing images from labels, inpainting, and many other purposes. We have benefited from this architecture in our methodology for various steps.

3.2 Objective Functions

As you may know, the Hausdorff Distance is widely used in evaluating medical image segmentation methods. In [Karimi and Salcudean 2019], they propose a novel loss function aiming at reducing Hausdorff Distance. Basically, this function measures the maximum distance of a predicted segmentation boundary to the nearest ground-truth edge pixel. However, it is computationally costly, therefore, they introduce some versions which do the trick faster with the approximation methods but, of course, lose some accuracy on the way. We have used the version which utilises morphological erosion in our binary mask generation networks so that we can have a balanced trade-off between the cost and accuracy.

There are various objective functions that can be utilized in order to improve the results of inpainting. In [Jiang et al. 2020], they show that their novel objective function enhances the results of generative models in a sense that it can focus on frequency components that are hard to synthesise. Specifically, it applies Discrete Fourier transforms, and then extracts amplitude and phase information from converted frequencies for both images. With such properties, a 2D plane is constructed to do L2 loss for back-propagation. On top of these, this objective function is complementary, meaning that it is supposed to be implemented along with others. We use this objective function along with our spatial ones to make our results more accurate and look more crisp. We use this objective function along with our spatial ones to make the results more accurate and look more crisp by focusing on hard-frequencies like edges and corners within tumors.

4 METHODOLOGY

Our own pipeline can be divided into two parts; namely mask generation and inpainting.

In the mask generation part we try to produce binary masks aiming to form pathologically realistic looking tumour shapes. After that, we try to convert this binary mask to multi-class mask with tumour label information embedded. We have two different approaches to generate such masks but their main idea is the same. We want to achieve simplification while preserving tumour information as much as possible. First one is based on concentric circles whereas the second uses morphological operations and is called the blob approach.

In the inpainting part, we experiment with different network architectures and objective functions to improve upon chosen baselines [Isola et al. 2016; Kim et al. 2020].

4.1 Concentric Approach to Mask Generation

It can be seen in the figure 2, we give the first segmenter the brain mask as well as our computed concentric circle. As for the tumour label segmenter, we feed the same concentric circle but with binary tumour shape. This means that both networks can be trained modularly.

4.1.1 Computing a concentric circle. To compute a concentric circle, 5 parameters are required in order to embed tumour label information (See Figure 3). We first find the centre coordinates with the help of a bounding box applied on the given mask. Then, for each three sub-region, we calculate the radius by finding the smallest

circle that can fit all the pixels in the Y axis. As a result we can create corresponding concentric circles like in the figure 4.

4.1.2 Value Noise. Due to already limited number samples, we had to come up with techniques to handle unseen data in order to have both accuracy and realism. Otherwise, as we will see in our experiments, networks come in short generalising properly. First modification is done on the values of concentric circles. We apply Gaussian noise on them. We will refer to this technique as “Value Noise” in our experiments.

4.1.3 Radius Noise. Second modification here is more unique than the previous one. We created a scheme to sample varying radii depending on the sizes of these given radii. Basically, the larger the radius, the greater its variance. We do it by scaling Gaussian Noise exponentially as it can be seen in the following equation.

Let r_i be the radius value of a tumor label i in a concentric circle such that $i \in \{Edema, Enhancing, Necrotic\}$.

$$r_i^{sampled} = r_i + \mathcal{N}(\mu, \sigma^2) \cdot (2^{0.15 \cdot r_i} - 1) \quad (1)$$

where μ is 0, and σ is 0.1. In the figure 5, effect of exponential scaling can be seen when radius values are randomly sampled 1000 times for each tumor label.

Visual example can be seen in the appendix A.1. We will be referring to this technique as “Radius Noise” in our experiments.

4.1.4 Ablation Components. We will be comparing previously mentioned noise techniques as well as different objective functions in our ablation studies where concentric circles are used. Ablation components are as follows

- Objective Functions
 - Cross-Entropy
 - Weighted Cross-Entropy
 - Dice Loss
 - Hausdorff Loss
- Noise Techniques
 - Value Noise (4.1.2)
 - Radius Noise (4.1.3)

The reason behind doing an ablation study is to find the best combination among the settings given above. For example, A weighted sum of Cross Entropy and Dice Loss; we can use this joint objective function to leverage the flexibility of Dice loss for class imbalance and at same time use cross-entropy for curve smoothing. On top of that, for example, we can add both noises to achieve maximum possible generalisation.

4.2 Blob Approach to Mask Generation

4.2.1 Blob shape reconstruction. As demonstrated in the baseline approach for mask reconstructions, we have extended our research by defining another version for masks. In order to create these we have extracted a binary version of the brain and the tumour segmentation. Furthermore, we have applied a mode filter (with a size of 13) to the tumour segmentation. In order to extend our investigation, we have added salt pepper noise to edges of this blob to create a variation and experiment its impact on reconstructing edges. The two variations can be seen in Figure 6.

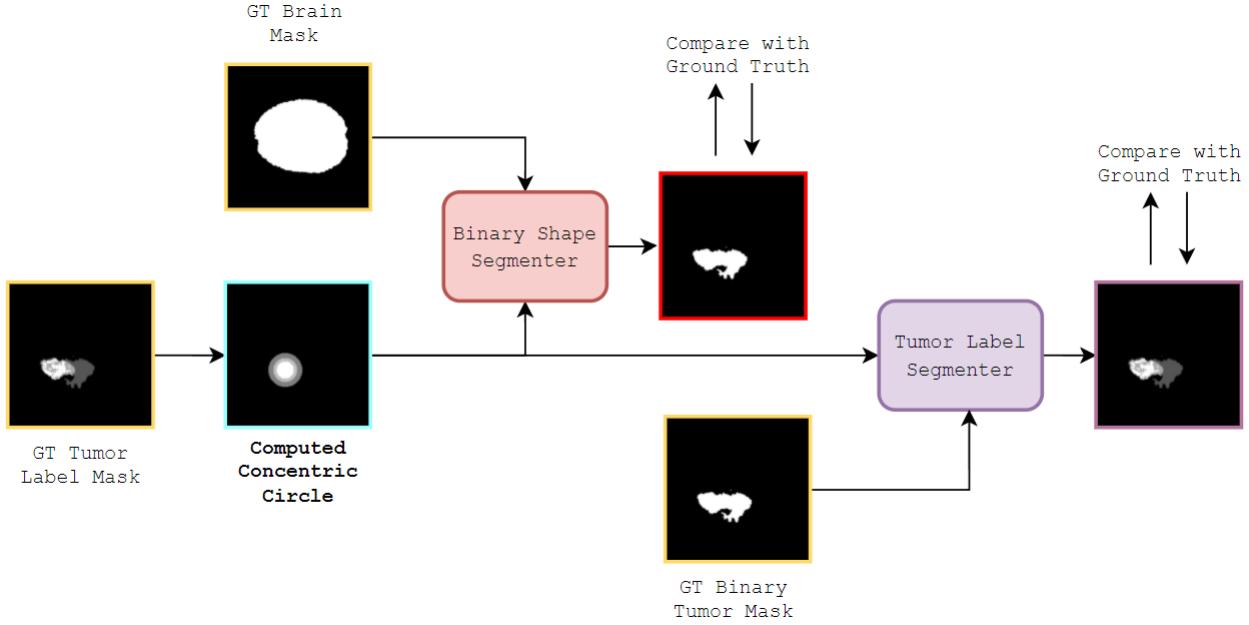


Fig. 2. Network flow of concentric approach to generate binary shape masks and multi-class tumor label masks. A concentric circle is computed based on ground truth tumor label mask. First, binary shape segmenter is used to generate realistic looking tumor masks, and then tumor label segmenter generates the multi-class version of this tumor mask so that sub-regions can be embedded and be used while inpainting. For Binary Shape and Tumor Label Segmenters, UNet [Ronneberger et al. 2015], UNet++ [Zhou et al. 2018] or [Kim et al. 2020] can be used as a network architecture.

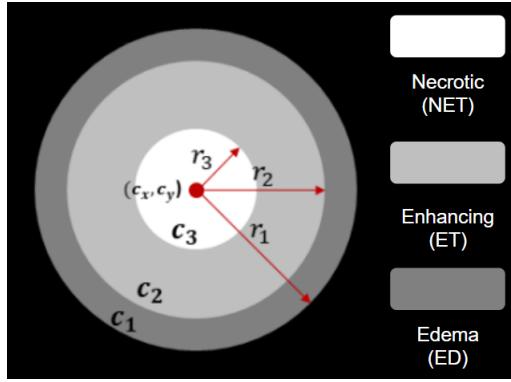


Fig. 3. Concentric circle and corresponding embedded tumor labels.

4.2.2 Tumor level segmentation. Another component that needs to be investigated is the level segmentation for tumor masks. As mentioned, pix2pix is an architecture that is applicable for segmentation map generation. Retrieving the binary representations of tumour segmentation and brain images a transition from a 1 level tumour image to a 3 level tumour image is attained as in Figure 7.

4.3 Tumor Inpainting

In our fusion-based baseline [Kim et al. 2020], we see that the authors decided to use a combination of loss functions for their generator network as follows

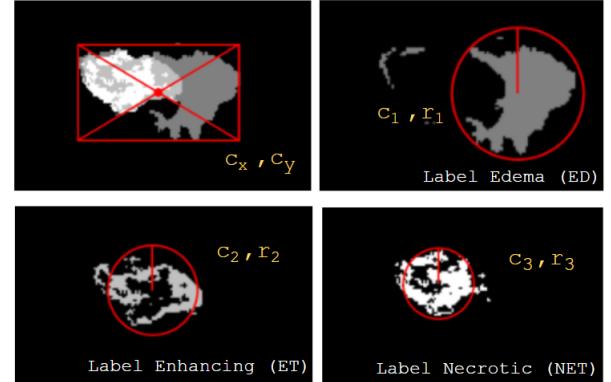


Fig. 4. In order to compute a concentric circle, 5 parameters are needed: c_x and c_y , center coordinates; r_1 , radius of a circle c_1 formed around label edema (ED); r_2 , radius of a circle c_2 formed around label enhancing (ET); r_3 , radius of a circle c_3 formed around label necrotic (NET).

$$\mathcal{L}_{baseline} = \mathcal{L}_{Adversarial} + \mathcal{L}_{GP} + \mathcal{L}_{LP} + \mathcal{L}_{Perceptual} \quad (2)$$

where \mathcal{L}_{LP} is the local (in-painted part) $L1$ pixel loss, \mathcal{L}_{GP} is the global $L1$ pixel loss, and $\mathcal{L}_{perceptual}$ is the $L1$ pixel loss based on outputs from the second layer's VGG-19 [Simonyan and Zisserman 2014]. One of the reasons why this baseline achieves competitive results is the perceptual loss they have included. This loss compares

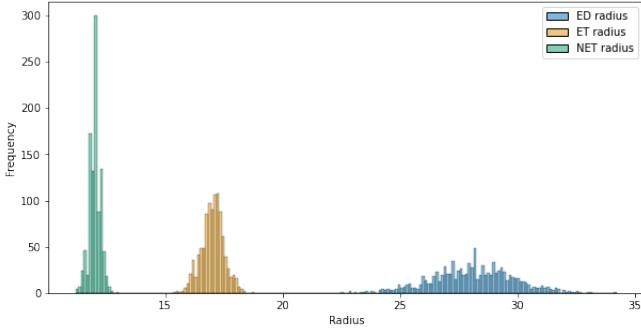


Fig. 5. 1000 random samples from an instance that had NET radius of 12, ET radius of 17, and ED radius of 28.

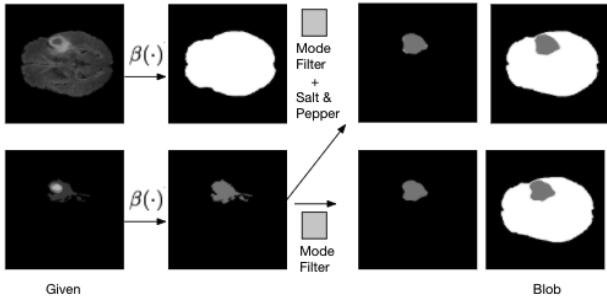


Fig. 6. Blob tumor generation

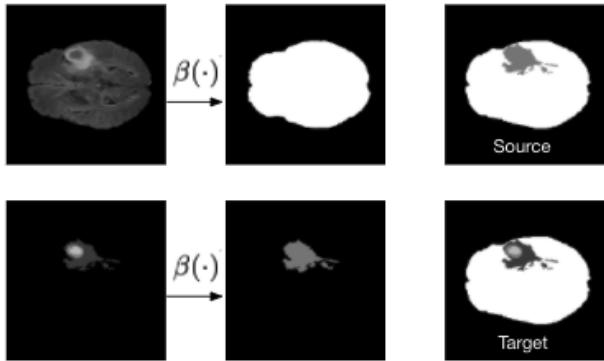


Fig. 7. Tumor level segmentation

the style in given images with the help of pre-trained VGG19. As a result, predictions look more detailed.

But we wanted to further improve the quality by using “Focal Frequency Loss” [Jiang et al. 2020] that we have shown you in our related work. This loss function has been able to improve the performance significantly for all modalities by being able to look more thoroughly into the high-frequency information like edges. We added this complementary loss as follows

$$\mathcal{L}_{\text{ours}} = \mathcal{L}_{\text{Adversarial}} + \mathcal{L}_{\text{GP}} + \mathcal{L}_{\text{LP}} + \mathcal{L}_{\text{Perceptual}} + w * \mathcal{L}_{\text{FFL}} \quad (3)$$

where w is 200 and \mathcal{L}_{FFL} is the Focal Frequency Loss.

4.4 Tumor Inpainting using alternative architectures

Another inpainting architecture used is pix2pix. This architecture is also experimented by the baseline approach. The architecture known as pix2pix uses combination of L1 and Binary Cross Entropy loss by default. Thus using comparative loss functions an extension is conducted to find a setup that would result in improvement in metrics.

4.5 Brain tissue deformation

In many cases where the tumor development is very advanced we can notice that brain tissue which surrounds the tumor is often deformed by the growing tumor mass. From our observations in brains that host large tumors often the large size of the tumor pushes the surrounding healthy brain parts and deforms the general shape of the brain. In an attempt to recreate this behaviour we experimented with an innovative method which uses unsupervised generating models. In an attempt to model this behavior we developed a Cycle-Gan that takes as input a healthy brain image as well as a real tumor shape mask (which we take from real brains with tumor) and produce an intermediate masked brain. In the intermediate step the tumor mask is applied on top of the healthy brain image. Then we insert this output into the Cycle-Gan model which generates a deformed brain image with the same tumor structure. The goal of the Cycle-Gan is to change the shape of the tissue that surrounds the tumor in a similar way to the deformations observed on real tumor images.

5 EXPERIMENTS

5.1 Dataset

We used the dataset called BraTS20 [Bakas et al. 2018]. There are total of 369 patient data in the BraTS20 and in each patient data includes 155 2D slices per modality (T1, T1ce, T2, Flair).

We had to do some preprocessing as we decided to only work on 2D slices. Shortly, we first normalized the slices within the patient data by dividing with the maximum recorded pixel value, and then discarded slices that had minimal information by removing the ones with less than 200 tumor pixels in their masks, in other words, we did not focus on really small tumours.

Furthermore, we had to reduce the number of samples for our experiments due to time and resource constraints; therefore, we decided to use only a quarter of all data. As a result: training set is reduced to 4.155 slices from 16.619; validation set is reduced to 531 slices from 2.125; test set is reduced to 531 slices from 2.125;

5.2 Evaluation Metrics

There are plenty of metrics to evaluate machine learning models. We have selected 11 of them based on the existing scientific literature. Dice, Precision, Recall are used for evaluation of the mask generation. Frechet inception distance (FID) and Image quality metrics (FSIM, SSIM, PSNR, RMSE, VIF, MSSSIM, GMSD) are used to evaluate synthetic images after inpainting. In the literature this is often advised to include many image quality metrics as possible since there is no golden rule. [Mason A 2020] note that Visual Information

Fidelity (VIF) and Feature-based Similarity Index (FSIM) are the most relevant image quality metrics to evaluate brain MRI. In fact, there are the two metrics with the highest correlation with radiologist evaluation. In the following sections we present experiment results with these metrics.

5.3 Concentric Approach

5.3.1 Binary Tumor Mask. We start with the ablation study on noise techniques and objective functions for concentric approach (See table 1). We observe that the one with the most best values is the “Value Noise” among those noise techniques; however, when we check these values column-wise, we see an exception that combined version performs the best specifically in Dice score. As for the objective functions. We stacked them up like joint objective functions. As a result, we can say that weighted combination of Cross-entropy, Dice loss and Hausdorff loss performs the best overall, especially when compared with the baseline [Kim et al. 2020].

As a trivial experiment (See table 2, we have implemented UNet++ [Zhou et al. 2018] to see if there can be improvement over different architecture. It is easy to say UNet++ is a better candidate because they have a similar number of learnable parameters yet the performance got boosted.

Here in the appendix B.2 we have some visuals on the validation set. As you might have noticed, at initial epochs, the network’s predictions look like shapes with not so complex boundaries. And, what if I told you that, at these initial epochs, these are the best results if we just consider the metric, precision for example. You can confirm it with the confusion matrix on the appendix B.3. It is really important to have several metrics as well as visual representations. Because, our sole goal is to produce realistic looking masks, and these initial predictions are definitely not. And to produce realistic looking shapes, one might say that the more over fitted the network is, the more realistic masks there will be but that’s counter productive in deep learning area. That’s the reason why we introduced noise techniques 4.1.2 and 4.1.3 to make our networks more robust in that sense.

5.3.2 Tumor Label Mask. Now we do the same ablation study for multi-class tumour label mask generation in the table 3. Difference here is that we have three classes now. We observe that the one with the most best values is the combined version of Radius and Value noise. We see that there is a significant performance increase when we apply our noise techniques, especially on Second and third labels, Enhancing and Necrotic parts which are more important than the first one, Edema. As for the objective functions they all perform better than the baseline; however, we can’t surely decide on the best.

We also wanted to see if the improvement still applies in this multi-class case when UNet++ [Zhou et al. 2018] is used. It surely does as you can see on the table 4.

Lastly, when we check the visuals in the appendix C.4 We notice that the realism problem in binary tumor generation does not exist but there are, of course, some hard cases as in the third column.

5.4 Blob Approach

5.4.1 Blob shape reconstruction. For the blob object reconstruction task, network architecture UNet was chosen and L1, MSE and SmoothL1 were used as the loss functions. The highest precision as false positives (false assigned tumor areas although they are not tumor) are very low as it was generated by applying a mode filter to the tumour segmentation. However the blob object itself has the lowest sensitivity as the false negatives are very high (in other terms the real tumors not being assigned to tumor)

Considering these quantitative results, SmoothL1 applied on pure blob definition has the best sensitivity and dice score among all as in Table 5. Also precision-wise the values are improved similar to the blob itself. L1 loss has created deformations in the edges of the blob object and this continues to become more severe as the training continued. However MSE and SmoothL1 were more responsive to the aim of the network.

Furthermore, adding saltpepper noise has helped the network to connect the blob object pieces and helped the tumor representation to extend, however the edges have become much more smoother than the ground truth.

5.4.2 Tumor level segmentation. For the tumor level segmentation we have first started experimenting with UNet architecture and used loss functions L1, MSE, SmoothL1 and their variations with reduction being summed.

Here in the quantitative results we can see that NET, has the highest precision and recall values as it was observed that the network was converging for core value easily. However the values for ED and ET the second and third layers of the tumor had a lower value as the areas towards the edges. With comparing the quantitative data in Table 6, we are able to observe that L1 variations using UNet architecture were more successful.

Furthermore, we have experimented on a conditional GAN (pix2pix) to generate the tumour segmentation level maps. Here the similar observations were found. Among the 6 networks compared, L1 with reduction equal to sum and Binary Cross Entropy with reduction sum was the outperforming network according to the quantitative results as in Table 7.

Binary Cross Entropy with logits combines a Sigmoid layer for defining the loss function Binary Cross Entropy with Logits Loss. To extend our previous experiment we have performed a new setup by applying Binary Cross Entropy with logits instead of Binary Cross Entropy. With this change L1 and Binary Cross Entropy with Logits Loss as well as Mean Squared Error and Binary Cross Entropy with Logits Loss have performed better comparatively as shown in Table 8.

Overall pix2pix did not add a major contribution to Unet when a global comparison is made.

5.5 Tumor Inpainting

5.5.1 Focal Frequency Loss. In the table 9 we have quantitative results of a trivial experiment on complementary objective function, that is Focal Frequency Loss. As you may remember from the methodology, we add this objective function on top of the fusion-based baseline’s [Kim et al. 2020] own objective function with some

Architecture	Objective	Noise	Dice \uparrow	Precision \uparrow	Sensitivity \uparrow
[Kim et al. 2020]	L1	-	0.7813	0.9060	0.9022
UNet	CE	-	0.8827	0.9054	0.9006
	CE	Radius	0.8680	0.8912	0.8814
	CE	Value	0.8811	0.9062	0.9024
	CE	Radius & Value	0.8692	0.8921	0.8824
UNet	$CE(w=1.0) + Dice(w=1.0)$	-	0.8830	0.9036	0.9022
	$CE(w=1.0) + Dice(w=1.0)$	Radius	0.8708	0.8909	0.8840
	$CE(w=1.0) + Dice(w=1.0)$	Value	0.8835	0.9037	0.9027
	$CE(w=1.0) + Dice(w=1.0)$	Radius & Value	0.8712	0.8897	0.8859
UNet	$CE(w=0.3) + Dice(w=0.7)$	-	0.8829	0.9039	0.9010
	$CE(w=0.3) + Dice(w=0.7)$	Radius	0.8712	0.8899	0.8843
	$CE(w=0.3) + Dice(w=0.7)$	Value	0.8834	0.9045	0.8891
	$CE(w=0.3) + Dice(w=0.7)$	Radius & Value	0.8709	0.8893	0.9014
UNet	$CE(w=0.25) + Dice(w=0.45) + Hausdorff(w=0.3)$	-	0.8827	0.9048	0.9008
	$CE(w=0.25) + Dice(w=0.45) + Hausdorff(w=0.3)$	Radius	0.8816	0.8884	0.8860
	$CE(w=0.25) + Dice(w=0.45) + Hausdorff(w=0.3)$	Value	0.8838	0.9050	0.8851
	$CE(w=0.25) + Dice(w=0.45) + Hausdorff(w=0.3)$	Radius & Value	0.8842	0.8909	0.9025

Table 1. Quantitative ablation study on noise techniques and objective functions in **binary shape mask generation networks**. Among those noise techniques, the one with the most best values is the “Value Noise”; however, combined version performs the best specifically in Dice score. Weighted combination of CE, Dice Loss and Hausdorff performs the best overall among objective functions. Compared with the baseline [Kim et al. 2020], there is a significant improvement over Dice score.

Architecture	Objective	Noise	Dice \uparrow	Precision \uparrow	Sensitivity \uparrow
UNet	$CE(w=0.3) + Dice(w=0.7)$	Radius	0.8712	0.8899	0.8843
UNet++	$CE(w=0.3) + Dice(w=0.7)$	Radius	0.8704	0.8864	0.8880
UNet	$CE(w=0.3) + Dice(w=0.7)$	Value	0.8834	0.9045	0.8891
UNet++	$CE(w=0.3) + Dice(w=0.7)$	Value	0.8836	0.9032	0.9035
UNet	$CE(w=0.3) + Dice(w=0.7)$	Radius & Value	0.8709	0.8893	0.9014
UNet++	$CE(w=0.3) + Dice(w=0.7)$	Radius & Value	0.8716	0.9011	0.9016
UNet	$CE(w=0.25) + Dice(w=0.45) + Hausdorff(w=0.3)$	Radius	0.8816	0.8884	0.8860
UNet++	$CE(w=0.25) + Dice(w=0.45) + Hausdorff(w=0.3)$	Radius	0.8812	0.8897	0.8851
UNet	$CE(w=0.25) + Dice(w=0.45) + Hausdorff(w=0.3)$	Value	0.8838	0.9050	0.8851
UNet++	$CE(w=0.25) + Dice(w=0.45) + Hausdorff(w=0.3)$	Value	0.8840	0.9031	0.9010
UNet	$CE(w=0.25) + Dice(w=0.45) + Hausdorff(w=0.3)$	Radius & Value	0.8842	0.8909	0.9025
UNet++	$CE(w=0.25) + Dice(w=0.45) + Hausdorff(w=0.3)$	Radius & Value	0.8847	0.8882	0.9030

Table 2. Quantitative results on UNet [Ronneberger et al. 2015] vs UNet++ [Zhou et al. 2018] in **binary shape mask generation networks**. UNet++ performs the best overall in both combination of objective functions.

weight. We can see that at epoch 50, it performs more or less competitively, but at epoch 100, tides change and we see significant improvement happening especially on the FID metric which we believe is the most eligible metric among others.

In the table 10, we use the same FID results but for each MR modality. As you can see they differ a lot. For example, the Flair version, the far right one, has the worst score yet we see the biggest improvement when we use the focal frequency loss.

Architecture	Objective	Noise	Dice Score ↑			Precision ↑			Sensitivity ↑		
			ED	ET	NET	ED	ET	NET	ED	ET	NET
[Kim et al. 2020]	L1	-	0.8110	0.3828	0.5507	0.8508	0.8386	0.4211	0.4542	0.6223	0.5578
UNet	CE	-	0.8224	0.4244	0.5661	0.8327	0.8642	0.4423	0.3937	0.578	0.5481
	CE	Radius	0.8163	0.4263	0.57	0.8347	0.8577	0.4426	0.3925	0.5906	0.5648
	CE	Value	0.8203	0.4167	0.5651	0.8346	0.8623	0.4399	0.3854	0.5823	0.5524
	CE	Radius & Value	0.8165	0.4277	0.5674	0.8358	0.8565	0.4373	0.3951	0.5868	0.5513
UNet	WCE	-	0.8214	0.4224	0.5609	0.8349	0.8643	0.4409	0.4036	0.5784	0.5395
	WCE	Radius	0.8175	0.4314	0.5708	0.838	0.8562	0.4276	0.3988	0.5823	0.5495
	WCE	Value	0.8200	0.4186	0.5612	0.8404	0.8548	0.4332	0.4242	0.5812	0.5464
	WCE	Radius & Value	0.8118	0.4217	0.5624	0.8362	0.8499	0.4358	0.3927	0.5702	0.5642
UNet	CE _(w=0.3) + Dice _(w=0.7)	-	0.8173	0.4539	0.5785	0.8359	0.854	0.4345	0.3771	0.5635	0.5757
	CE _(w=0.3) + Dice _(w=0.7)	Radius	0.8137	0.4619	0.5735	0.8312	0.8571	0.4214	0.3713	0.5673	0.5556
	CE _(w=0.3) + Dice _(w=0.7)	Value	0.8142	0.4484	0.5677	0.8328	0.8521	0.4192	0.3773	0.5592	0.5522
	CE _(w=0.3) + Dice _(w=0.7)	Radius & Value	0.8183	0.4609	0.5724	0.8363	0.8514	0.428	0.3895	0.5602	0.5591
UNet	CE _(w=1.0) + Dice _(w=1.0)	-	0.8171	0.4779	0.5758	0.8277	0.8582	0.4408	0.3589	0.565	0.5766
	CE _(w=1.0) + Dice _(w=1.0)	Radius	0.8144	0.4473	0.5725	0.8379	0.8465	0.4227	0.3734	0.5596	0.5689
	CE _(w=1.0) + Dice _(w=1.0)	Value	0.8153	0.4865	0.5822	0.835	0.8517	0.441	0.3947	0.5701	0.5709
	CE _(w=1.0) + Dice _(w=1.0)	Radius & Value	0.8177	0.4896	0.5737	0.8342	0.8578	0.4458	0.3781	0.5641	0.5682
UNet	WCE _(w=1.0) + Dice _(w=1.0)	-	0.8156	0.46	0.5646	0.826	0.8627	0.4225	0.3848	0.5731	0.5281
	WCE _(w=1.0) + Dice _(w=1.0)	Radius	0.8134	0.4698	0.5742	0.8394	0.8494	0.4335	0.4026	0.5683	0.5582
	WCE _(w=1.0) + Dice _(w=1.0)	Value	0.8143	0.4729	0.5733	0.8255	0.8557	0.4376	0.3658	0.5603	0.5653
	WCE _(w=1.0) + Dice _(w=1.0)	Radius & Value	0.8091	0.4823	0.5837	0.8409	0.8406	0.4309	0.3856	0.5586	0.5868

Table 3. Quantitative ablation study on noise techniques and objective functions in **tumor label mask generation networks**. Among those noise techniques, the one with the most best values is the “Value Noise”; however, combined version performs the best specifically in Dice score. Weighted combination of CE, Dice Loss and Hausdorff performs the best overall among objective functions. Compared with the baseline [Kim et al. 2020], there is a significant improvement over Dice score.

Architecture	Objective	Noise	Dice Score ↑			Precision ↑			Sensitivity ↑		
			ED	ET	NET	ED	ET	NET	ED	ET	NET
UNet	CE _(w=0.3) + Dice _(w=0.7)	Radius	0.8137	0.4619	0.5735	0.8312	0.8571	0.4214	0.3713	0.5673	0.5556
UNet++	CE _(w=0.3) + Dice _(w=0.7)	Radius	0.8162	0.4596	0.5754	0.8349	0.8514	0.4342	0.3829	0.5553	0.5627
UNet	CE _(w=0.3) + Dice _(w=0.7)	Value	0.8142	0.4484	0.5677	0.8328	0.8521	0.4192	0.3773	0.5592	0.5522
UNet++	CE _(w=0.3) + Dice _(w=0.7)	Value	0.8169	0.4781	0.5770	0.8308	0.8545	0.4411	0.3752	0.5563	0.5681
UNet	CE _(w=0.3) + Dice _(w=0.7)	Radius & Value	0.8183	0.4609	0.5724	0.8363	0.8514	0.428	0.3895	0.5602	0.5591
UNet++	CE _(w=0.3) + Dice _(w=0.7)	Radius & Value	0.8192	0.4715	0.5769	0.8452	0.8519	0.4489	0.4154	0.5718	0.5653

Table 4. Quantitative results on UNet [Ronneberger et al. 2015] vs UNet++ [Zhou et al. 2018] in **tumor label mask generation networks**. UNet++ performs the best overall in both combination of objective functions.

In the figure 9b, we show the magnitude spectrums for single Flair MRI instance. It might be a little bit hard to see the difference between them in (a); hence in (b), contrast is increased to ease that problem. Now you might be able to notice the colour distributions. FFL works as intended since there seems to be notable changes in frequency domain when we look at these.

We said that FFL works on hard cases, hard frequencies such as edges and corners. We can actually observe that in the figure 9. Compare the inside of the tumor areas on the right with the ground truth version on the left. Without FFL, the inpainting has more blurriness whereas with FFL used, there are formation of

boundaries defined within. As a result, this is a good step towards achieving realistic results.

5.2 Pix2Pix Architecture. We have further experimented with Pix2Pix architecture for the tumor inpainting task. Here we have experimented with a combination of L1, Mean Squared Error and SmoothL1 losses being averaged and summed and observed the changes in the network quantitative results by using Binary Cross Entropy and Binary Cross Entropy with logis loss. We have also introduced SSIM loss and added to L1 and Mean Squared Error for 2 other variations. Below are some interpretations of the output which is provided at Table 11

Architecture	Objective	Noise	Dice ↑	Precision ↑	Sensitivity ↑
User Guided Input	-	-	0.9080	0.9484	0.8784
UNet	L1	ModeFilter	0.9309	0.9439	0.9216
UNet	MSE	ModeFilter	0.9321	0.9420	0.9255
UNet	SmoothL1	ModeFilter	0.9327	0.9420	0.9277
UNet	MSE	ModeFilter + Salt & Pepper	0.9249	0.9315	0.9215
UNet	SmoothL1	ModeFilter + Salt & Pepper	0.9245	0.9361	0.9163

Table 5. Quantitative results on using ground truth mask value to generate a simpler representation as blob object and further reconstructing using GAN with a UNet architecture

Architecture	Objective	Dice Score ↑			Precision ↑			Sensitivity ↑		
		ED	ET	NET	ED	ET	NET	ED	ET	NET
UNet	L1	0.2237	0.1618	0.7825	0.5452	0.2338	0.7582	0.2220	0.2496	0.8943
UNet	MSE	0.1461	0.2332	0.7296	0.5107	0.2189	0.8304	0.1363	0.5567	0.7269
UNet	SmoothL1	0.0875	0.2202	0.7555	0.5575	0.2263	0.8002	0.0720	0.5044	0.7939
UNet	L1 (reduction=sum)	0.2169	0.1732	0.7869	0.5593	0.2429	0.7607	0.2150	0.2675	0.8935
UNet	MSE (reduction=sum)	0.1107	0.2370	0.7223	0.4168	0.2151	0.8221	0.0990	0.5879	0.7224
UNet	SmoothL1 (reduction=sum)	0.0611	0.2180	0.7217	0.3694	0.2020	0.8157	0.0567	0.5715	0.7263

Table 6. Quantitative data comparison between various loss functions for Unet architecture for tumor level segmentation task

Architecture	Objective	Dice Score ↑			Precision ↑			Sensitivity ↑		
		ED	ET	NET	ED	ET	NET	ED	ET	NET
Pix2Pix	L1 + BCE	0.1391	0.1644	0.7553	0.3291	0.2126	0.7508	0.1521	0.3152	0.8510
Pix2Pix	MSE + BCE	0.1846	0.1172	0.6737	0.2435	0.1844	0.7350	0.2932	0.2052	0.7274
Pix2Pix	SmoothL1 + BCE	0.2070	0.0868	0.7151	0.2429	0.1690	0.7274	0.3268	0.1394	0.7844
Pix2Pix	L1 (reduction=sum) + BCE (reduction=sum)	0.2007	0.1900	0.7725	0.5155	0.2273	0.7739	0.2144	0.3252	0.8561
Pix2Pix	MSE (reduction=sum) + BCE (reduction=sum)	0.1020	0.2369	0.7404	0.4713	0.2299	0.8188	0.0999	0.5518	0.7558
Pix2Pix	SmoothL1 (reduction=sum) + BCE (reduction=sum)	0.0646	0.2341	0.7208	0.3950	0.2201	0.8356	0.0621	0.6146	0.7104

Table 7. Quantitative data comparison between various loss functions for Pix2Pix architecture for tumor level segmentation task

Architecture	Objective	Dice Score ↑			Precision ↑			Sensitivity ↑		
		ED	ET	NET	ED	ET	NET	ED	ET	NET
Pix2Pix	L1 + BCE Logits Loss	0.2242	0.1657	0.7791	0.5107	0.2393	0.7690	0.2471	0.2659	0.8715
Pix2Pix	MSE + BCE Logits Loss	0.0853	0.2261	0.7283	0.4302	0.2052	0.8261	0.0753	0.5885	0.7277
Pix2Pix	SmoothL1 + BCE Logits Loss	0.0843	0.1995	0.7757	0.4675	0.2247	0.7825	0.0694	0.4461	0.8465
Pix2Pix	L1 (reduction=sum) + BCE Logits Loss (reduction=sum)	0.1929	0.2005	0.7739	0.4891	0.2540	0.7732	0.2000	0.3514	0.8573
Pix2Pix	MSE (reduction=sum) + BCE Logits Loss (reduction=sum)	0.1143	0.2225	0.7500	0.4787	0.2186	0.8119	0.1139	0.5016	0.7757
Pix2Pix	SmoothL1 (reduction=sum) + BCE Logits Loss (reduction=sum)	0.0463	0.2155	0.7618	0.3444	0.2327	0.7883	0.0345	0.4869	0.8216

Table 8. Quantitative data comparison between various loss functions for Pix2Pix architecture for tumor level segmentation task, introducing Logits Loss

Comparing top 3 networks we can see that the best metrics were hold by L1 and Binary Cross Entropy. When we compare globally all metrics, L1 summed combined SSIM and Binary Cross Entropy with Logits Loss summed ranks highest.

In terms of FID value globally, L1 and Binary Cross Entropy with Logits Loss is the network lowest. It is important to emphasise

especially on the FID metric which we believe is the most eligible metric among others.

L1 and Binary Cross Entropy with logits loss, the one with the lowest FID score, preserves more of the structure.

Experiment (25% Training Data)	$\text{FID}_{avg} \downarrow$	$\text{SSIM}_{avg} \uparrow$	$\text{MS-SSIM}_{avg} \uparrow$	$\text{GMSD}_{avg} \downarrow$	$\text{PSNR}_{avg} \uparrow$	$\text{VIF}_{avg} \uparrow$	$\text{RMSE}_{avg} \downarrow$
[Kim et al. 2020] at epoch 50	12.6024	0.9846	0.9852	0.0390	38.7301	0.8929	0.0072
[Kim et al. 2020] w/ FFL at epoch 50	12.7716	0.9845	0.9850	0.0395	38.6805	0.8927	0.0073
[Kim et al. 2020] at epoch 100	11.0527	0.9838	0.9845	0.0387	38.3602	0.8865	0.0075
[Kim et al. 2020] w/ FFL at epoch 100	10.8257	0.9837	0.9846	0.0391	38.3816	0.8867	0.0073

Table 9. Quantitative results on [Kim et al. 2020] with Focal Frequency Loss added. FFL starts showing its power at a later stage of the training. Also, FID metric makes the comparison more eligible as other metrics don't give much information when different epochs are considered.

Experiment (25% Training Data)	Modalities				
	$\text{FID}_{avg} \downarrow$	$\text{FID}_{T1} \downarrow$	$\text{FID}_{T1c} \downarrow$	$\text{FID}_{T2} \downarrow$	$\text{FID}_{FLAIR} \downarrow$
[Kim et al. 2020] at epoch 50	12.6024	13.2748	13.0540	8.1189	15.9619
[Kim et al. 2020] w/ FFL at epoch 50	12.7716	12.8392	13.0632	8.4568	16.7275
[Kim et al. 2020] at epoch 100	11.0527	10.9617	11.1145	7.5041	14.6308
[Kim et al. 2020] w/ FFL at epoch 100	10.8257	10.5596	11.1637	7.7615	13.8182

Table 10. Same inpainting results on FID metric from the table 9 but for each MR modality. The results differ a lot modality-wise. For example, the Flair version, has the worst score yet it has seen the biggest improvement when focal frequency loss is used.

Objective	$\text{FID}_{avg} \downarrow$	$\text{SSIM}_{avg} \uparrow$	$\text{MS-SSIM}_{avg} \uparrow$	$\text{PSNR}_{avg} \uparrow$	$\text{VIF}_{avg} \uparrow$	$\text{RMSE}_{avg} \downarrow$
L1 + BCE	20.5172	0.9775	0.9809	34.7473	0.8189	0.0082
MSE + BCE	64.0382	0.8627	0.9748	32.1564	0.6652	0.0103
SmoothL1 + BCE	97.2526	0.9287	0.9601	29.4557	0.4479	0.0139
L1 + BCE Logits Loss	14.9656	0.9799	0.9821	35.8716	0.8379	0.0076
MSE + BCE Logits Loss	17.6472	0.9791	0.9826	35.4395	0.7796	0.0077
SmoothL1 + BCE Logits Loss	22.9451	0.9759	0.9798	34.2863	0.7215	0.0083
L1 & SSIM + BCE Logits Loss	16.5013	0.9796	0.9821	35.4624	0.8248	0.0077
MSE & SSIM + BCE Logits Loss	18.2644	0.9791	0.9824	35.2186	0.7726	0.0077
L1 (reduction=sum) + BCE Logits Loss (reduction=sum)	17.4435	0.9816	0.9838	36.4325	0.8511	0.0071
MSE (reduction=sum) + BCE Logits Loss (reduction=sum)	17.4378	0.9811	0.9835	36.1798	0.8304	0.0072
SmoothL1 (reduction=sum) + BCE Logits Loss (reduction=sum)	17.0057	0.9808	0.9834	35.9236	0.8251	0.0073
L1 (reduction=sum) & SSIM + BCE Logits Loss (reduction=sum)	17.4516	0.9819	0.9842	36.5252	0.8489	0.0069
MSE (reduction=sum) & SSIM + BCE Logits Loss (reduction=sum)	17.1549	0.9810	0.9836	36.1934	0.8284	0.0072

Table 11. Quantitative results using Pix2Pix as an Inpainting architecture

6 DISCUSSION

6.1 MRI synthesis

From the pix2pix baseline [Isola et al. 2016] we conducted series of experiments to go from segmentation to MRI images. For all tests we used the same settings: 25% of the BraTS20 dataset as mentioned in Section 5.1, and 10 epochs (due to computation limitation). Table 12 presents comparisons with the baseline. Each line corresponds to a new setting, the parameter which was changed is written for each of them. Of course, we are not specialists, but with these experiments we tried to qualitatively analyze the metrics.

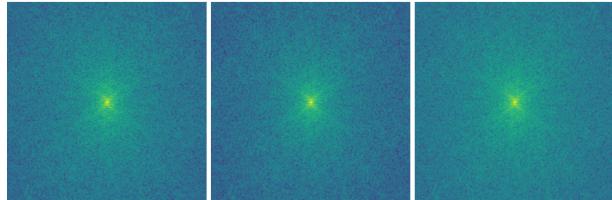
Here are some qualitative analysis:

First we consider Figure 11a, which corresponds to the best FSIM. The image is blurred but the shape details are high. For instance, details of tumor boundaries are high. This is coherent with the fact that FSIM measure's first feature is phase congruency, which seem to reflect the edge of the image. However, visually having only good edges like this, is not satisfying.

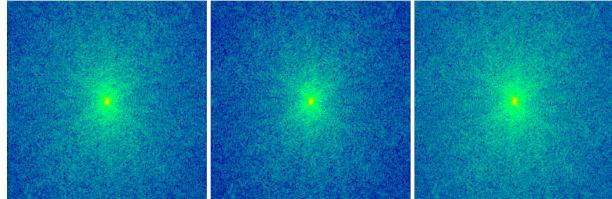
We then consider Figure 11b and Figure 11c, which correspond to the two models with best VIF. Whereas the previous example, the brain structure is present. However, we note that the tumor is blurred again and that no focus is given to it. The VIF reflects good overall visuals in our experiments.

Table 12. Experiments' settings and corresponding metrics for MRI synthesis

Experiment (25% Training Data)	FID ↓	SSIM ↑	MS-SSIM ↑	GMSD ↓	PSNR ↑	VIF ↑	RMSE ↓	FSIM ↑
Pix2pix_baseline	49,58	0,654	0,21983	0,45577	39,94	10,525	0,01042	0,289
Pix2pix_BCE-wl-Loss	53,87	0,6313	0,23	0,4466	39,22	13,46	0,01108	0,4861
Pix2pix_MSELoss	68,99	0,5881	0,1867	0,4769	38,19	7,499	0,01251	0,234
Pix2pix_SmoothL1Loss	42,78	0,598	0,23099	0,471	40,44	8,79	0,00973	0,25
Pix2pix_D_RMSProp	61,43	0,6307	0,19405	0,46355	38,78	9,340	0,01178	0,2639
Pix2pix_G_RMSProp	75,26	0,6645	0,18209	0,45308	37,98	8,717	0,0129	0,2720
Pix2pix_MSELoss_D-lr-0.00002	37,22	0,5865	0,23942	0,47629	41,13	12,027	0,00906	0,2539
Pix2pix_D-lr-0.00002	58,72	0,6092	0,20352	0,4679	39,18	9,956	0,01133	0,2658
Pix2pix_G-lr-0,0005	54,17	0,6290	0,21263	0,4628	39,39	9,930	0,01097	0,2780
Pix2pix_MSELoss_λ=32	37,23	0,597	0,2117	0,47319	39,51	10,32	0,01102	0,2617
Pix2pix_λ=1000	55,77	0,6282	0,2156	0,46439	39,20	10,409	0,01117	0,2739
Pix2pix_λ=10	56,28	0,6423	0,21047	0,45879	39,45	7,88	0,01109	0,2753



(a) Original: Ground truth, without FFL and with FFL



(b) Contrast increased: Ground truth, without FFL and with FFL

Fig. 8. Magnitude spectrums visuals retrieved from the experiment showed in the table 10 averaged over 32 Flair MRI at 100th epoch. With FFL used, color distribution looks similar to the ground truth when contrast levels are increased

When we look at Figure 11d corresponding to the model with the highest number of best values, a compromise appear. In fact, there are tumor boundaries, there are different texture, and the overall structure of the brain is present.

In short, compared to inpainting methods presented before, no focus is given on the tumor: there are often no clear tumor boundaries. A good illustration of this absence of tumor focus is given by Figure 11e, the tumor is only visible because of the color change. We can note that there are more pixels for the brain than for the tumor, therefore the metrics can't access this tumor focus. One strength of inpainting methods is to focus on the anomaly: the tumor.

7 CONCLUSIONS

As seen in the binary tumour mask generation, there is a trade-off between accuracy and realism. We believe we have made good improvements on that by combining objective functions and adding specific noises. As for the multi-class tumour label mask generation, we believe we are in a good spot with both concentric and blob approaches, they just need to be trained with the whole data so that we see the full potential. One other good thing about these approaches is that we can synthesise brain MRI with tumour inside really easily by just playing with few parameters.

We achieved competitive results in inpainting with different objective functions, noise techniques and architectures. Focal Frequency Loss performed really well in our preliminary inpainting experiments. Since it has many variables to play with, we see many opportunities.

8 FUTURE WORK

Data Augmentation. We had this idea of bench-marking via using our inpainting methods as a data augmentation tool. With this way, it would be a really good way to show if our synthetic tumour data can improve the performances when used as a dataset. Some of the experiment ideas are

- Model(s) trained with **real dataset** vs Model(s) trained with **synthetic dataset**
- Model(s) trained with **real dataset** vs Model(s) trained with **real + synthetic dataset**

Edge-based Objective Functions. For generative networks tasked with inpainting, edge-based objective functions which use operations like Sobel and Canny can be explored in order to achieve more realistic results.

Small Tumors. In our experiments we didn't take small tumours into account but It would be a really interesting experiment to use them. Synthesising small but expanding tumour would probably come in handy in the industry.

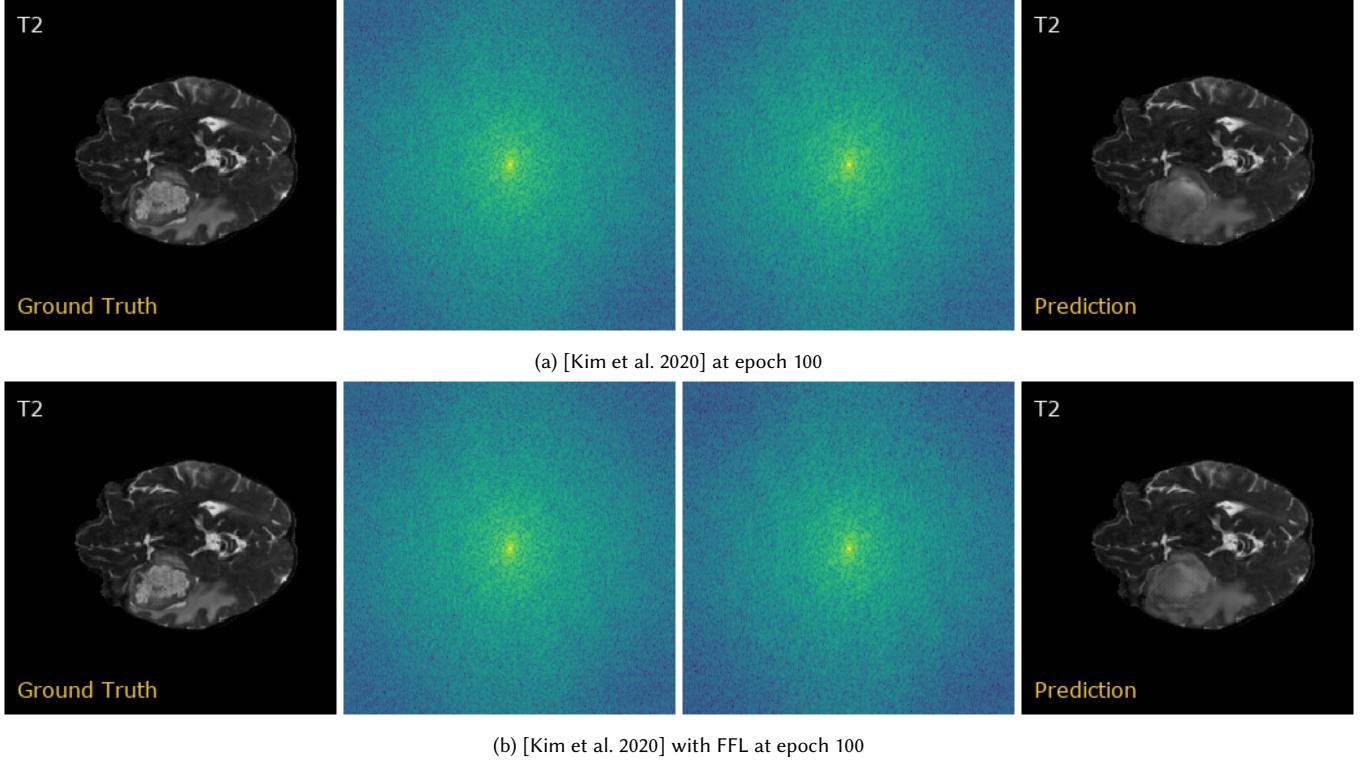
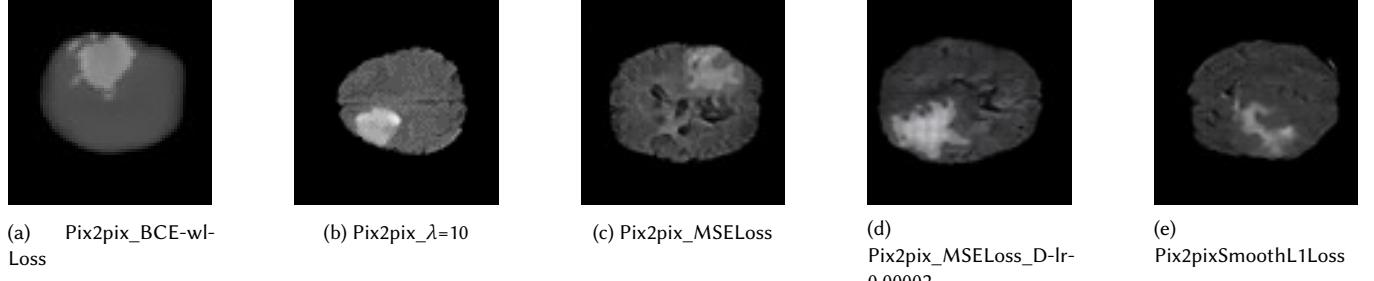


Fig. 9. FFL works on hard cases, hard frequencies such as edges and corners. Here it can be observed within the in-painted tumor area. With FFL (bottom-right), there are formation of boundaries resembling the ground truth on the left

Fig. 10. Synthetic MRI for different parameter settings



Different Inputs. We thought about using different kinds of input to make our predictions more realistic. For example, check the figure 2, we can feed the binary tumor mask network with Brain ATLAS information [Iqbal et al. 2018] instead of binary brain mask so that the network can generate shapes that are realistic according to the parts in the atlas.

Speckles. We conducted experiments focused on speckles. First, as a preprocessing step before input the image into a network, this step is called speckle removal [Bin-Habtoor and Al-amri 2016]. Concretely, we used image filtering with mean, median and crimmins

filter. Second, as a post-processing step we conducted speckle generation [Bargsten and Schlaefler 2020; Goodman 2007]. We implemented this process with series of Fourier transforms.

$$\begin{aligned}
 I_{\text{speckles}}(x, y) &= |F^{-1}\{F\{I_{\text{source}}(x, y) \cdot e^{j\varphi(x, y)}\} \times \text{rect}_d(x, y)\}| \\
 I_{\text{speckles}}(x, y) &= |I_{\text{source}}(x, y) \cdot e^{j\varphi(x, y)} * F^{-1}\{\text{rect}_d(x, y)\}| \\
 I_{\text{speckles}}(x, y) &= |I_{\text{source}}(x, y) \cdot e^{j\varphi(x, y)} * \text{sinc}_d(x, y)|
 \end{aligned} \tag{4}$$

We didn't continue in depth experiments during our project. In fact, we focused on MRI and MRI doesn't contain speckles. However, in optic-based medical imagery (for instance in CT imagery or

Inpainting in Medical Imaging

ultrasound imagery) speckles are present. Therefore, in future work using optic-based medical imagery, we think that speckles can be interesting to consider.

ACKNOWLEDGMENTS

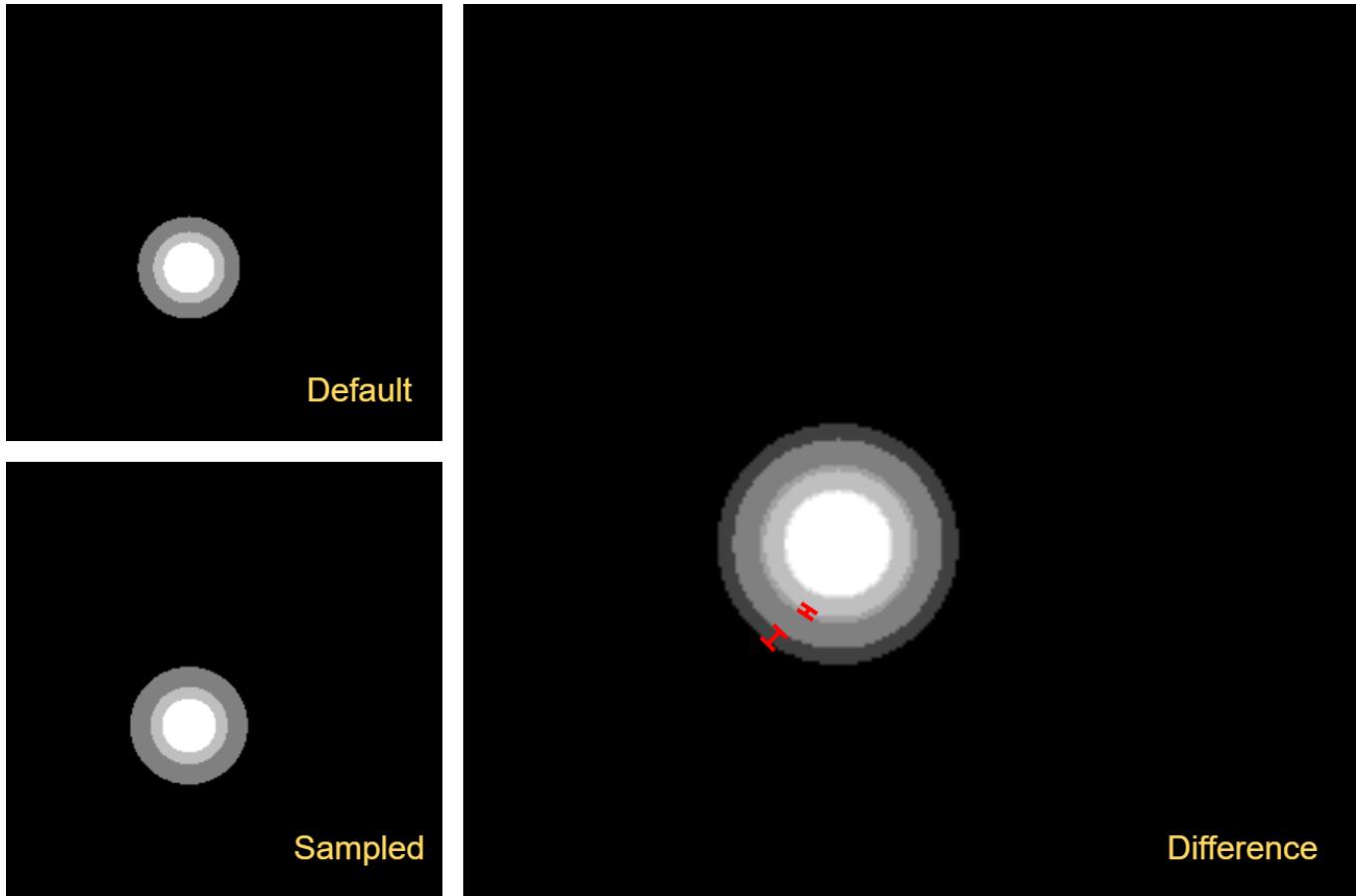
We would like to thank Yousef Yeganeh and Azade Farshad for providing great assistance in our project.

REFERENCES

- Spyridon Bakas, Mauricio Reyes, and Andras Jakab. 2018. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. *arXiv preprint arXiv: Arxiv-1811.02629* (2018).
- Lennart Bargsten and Alexander Schlaefer. 2020. SpeckleGAN: a generative adversarial network with an adaptive speckle layer to augment limited training data for ultrasound image processing. *International Journal of Computer Assisted Radiology and Surgery* 15 (06 2020). DOI : <https://doi.org/10.1007/s11548-020-02203-1>
- Abdulaziz Saleh Yeslem Bin-Habtoor and Salem Saleh Al-amri. 2016. Removal Speckle Noise from Medical Image Using Image Processing Techniques. *International Journal of Computer Science and Information Technologies* 7(1) (2016), 375–377.
- J.W. Goodman. 2007. *Speckle Phenomena in Optics: Theory and Applications*. Roberts & Company. <https://books.google.de/books?id=TynXEcS0DncC>
- Asim Iqbal, Romesh Khan, and Theofanis Karayannis. 2018. Developing Brain Atlas through Deep Learning. *arXiv preprint arXiv: Arxiv-1807.03440* (2018).
- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2016. Image-to-Image Translation with Conditional Adversarial Networks. *arXiv preprint arXiv: Arxiv-1611.07004* (2016).
- Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. 2020. Focal Frequency Loss for Image Reconstruction and Synthesis. *arXiv preprint arXiv: Arxiv-2012.12821* (2020).
- Davood Karimi and Septimiu E. Salcudean. 2019. Reducing the Hausdorff Distance in Medical Image Segmentation with Convolutional Neural Networks. *arXiv preprint arXiv: Arxiv-1904.10030* (2019).
- Sunho Kim, Byungjai Kim, and HyunWook Park. 2020. Synthesis of Brain Tumor MR Images for Learning Data Augmentation. *arXiv preprint arXiv: Arxiv-2003.07526* (2020).
- Clarke SE Costa A Schmidt M Keough V Huynh T Beyea S. Mason A, Rioux J. 2020. Comparison of Objective Image Quality Metrics to Expert Radiologists' Scoring of Diagnostic Quality of MR Images. *IEEE Trans Med Imaging* 39(4) (2020).
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv preprint arXiv: Arxiv-1505.04597* (2015).
- Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv: Arxiv-1409.1556* (2014).
- Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. 2018. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *arXiv preprint arXiv: Arxiv-1807.10165* (2018).

APPENDIX

A.1 A concentric circle with sampled radiiuses



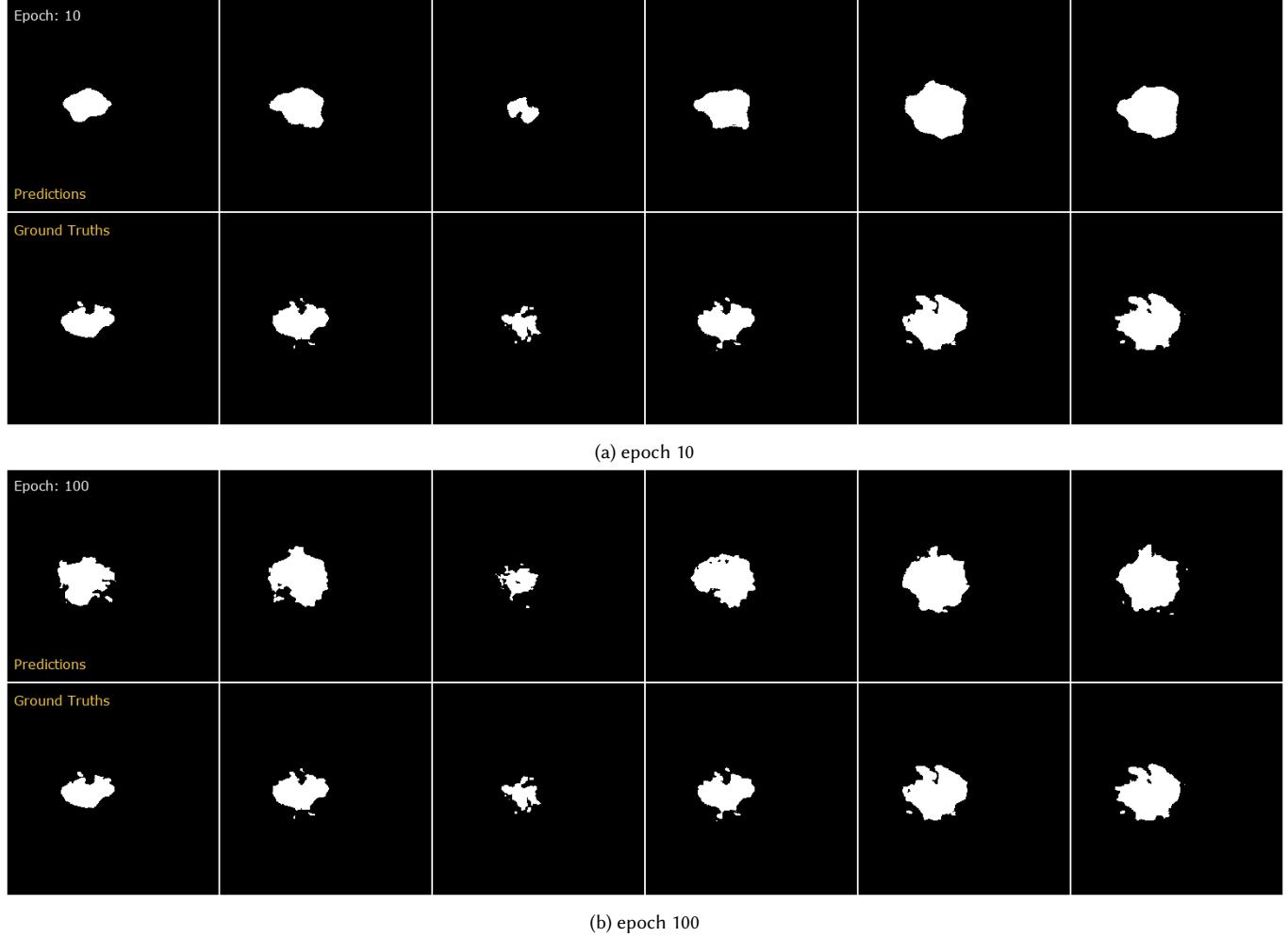
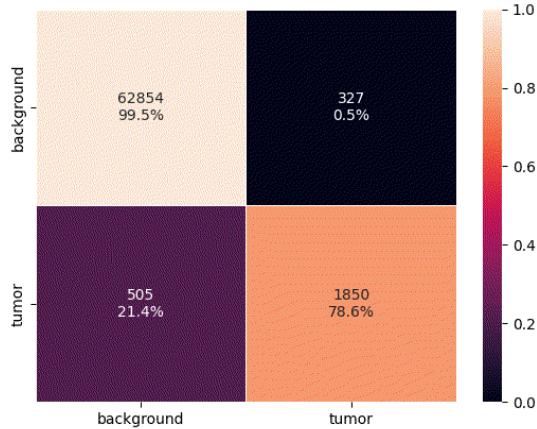


Fig. 12. UNET_{(CE_(w=0.25)+Dice_(w=0.45)+Hausdorff_(w=0.3))}, Visual Validation Results for Binary Tumor Mask Generation.

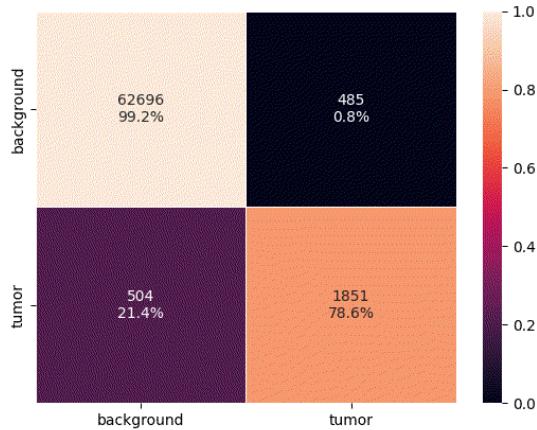
B.2 Visual Validation Results for Binary Tumor Mask Generation

Epoch: 10



(a) epoch 10

Epoch: 100



(b) epoch 100

Fig. 13. UNET_(CE_{w=0.25}+Dice_{w=0.45}+Hausdorff_{w=0.3}), Validation Confusion Matrix for Binary Tumor Mask Generation.

B.3 Validation Confusion Matrix for Binary Tumor Mask Generation

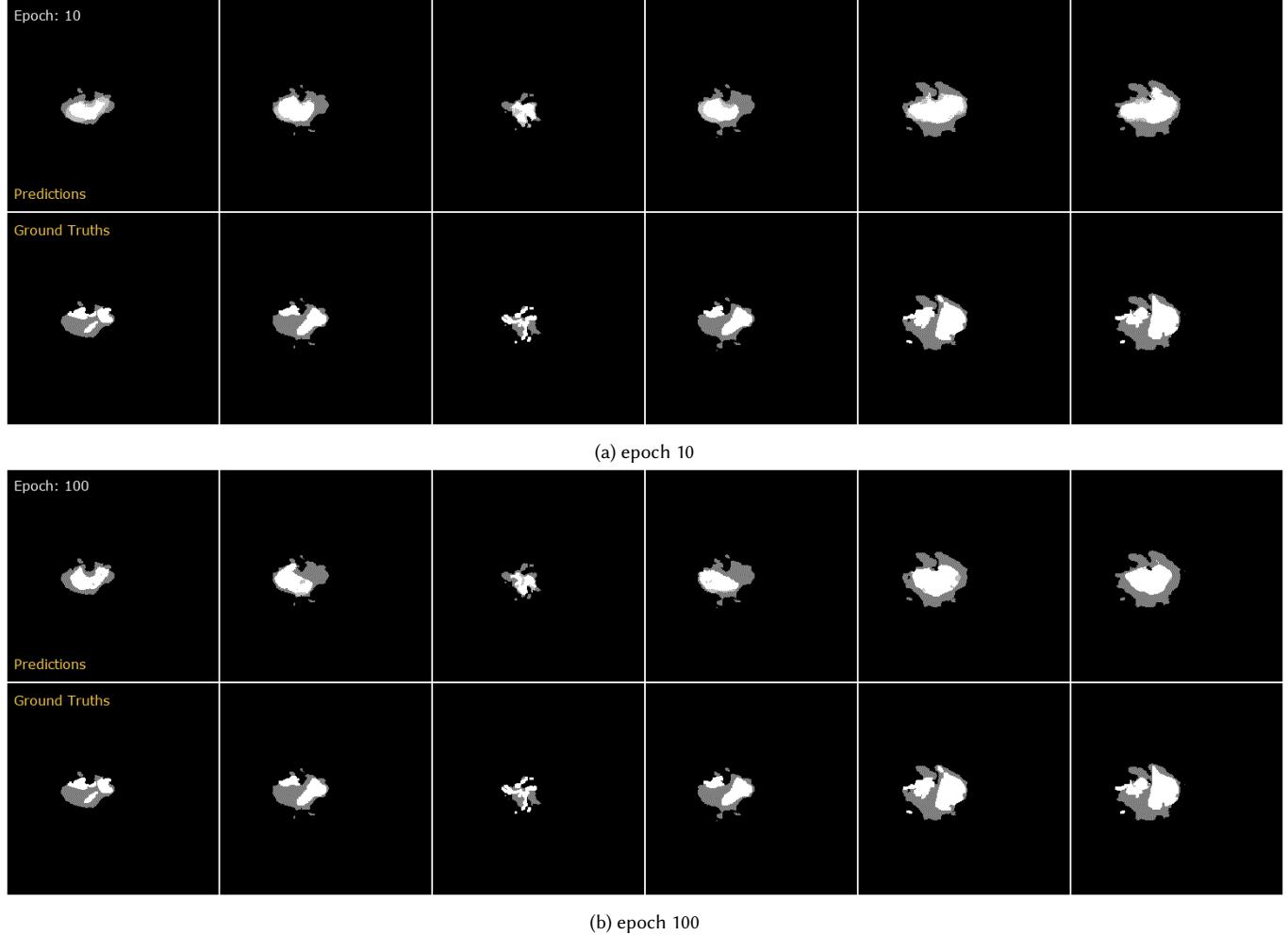
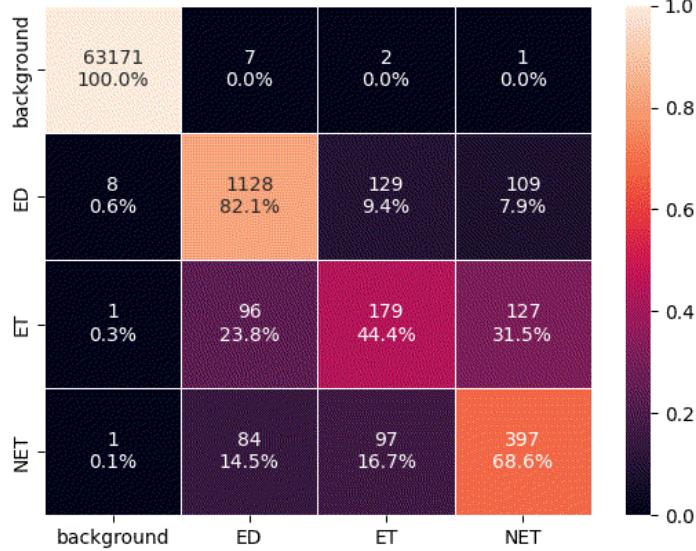


Fig. 14. UNET_{(CE_(w=0.3)+Dice_(w=0.7))}, Visual Validation Results for Tumor Label Mask Generation.

C.4 Visual Validation Results for Tumor Label Mask Generation

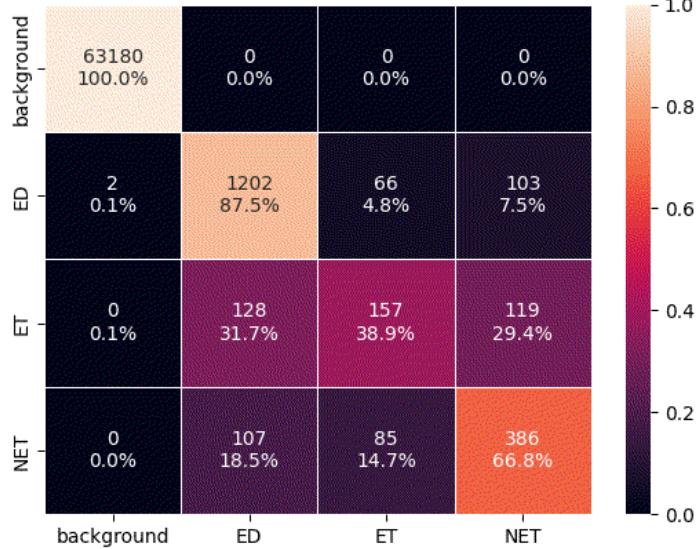
Inpainting in Medical Imaging

Epoch: 10



(a) epoch 10

Epoch: 100



(b) epoch 100

Fig. 15. UNET_(CE_{w=0.3}+Dice_{w=0.7}), Validation Confusion Matrix for Tumor Label Mask Generation. **Note:** values are very erratic per epoch, hence dual comparisons might not be ideal.

C.5 Validation Confusion Matrix for Tumor Label Mask Generation