Dose Prediction in Radiotherapy

Mona Mokart MONA.MOKART@TELECOM-PARIS.FR
Théau Blanchard THEAU.BLANCHARD@TELECOM-PARIS.FR
Codalab team : tblanchard *Télécom Paris / ENS Paris-Saclay*

Abstract

Dose prediction is an important procedure in radiotherapy that helps to guide the treatment. Yet determining precise dose comes at the cost of very expensive simulations. Here we present a deep-learning based approach tackling the problem of dose prediction using CT scans of the patient. All code is available on our github.

Keywords: deep learning, medical imaging, dose prediction, supervised learning, unet, domain transfer

1. Introduction

As the accuracy of the radiotherapy tools improves, so should the treatement of the patient. More particularly, having a good treatement plan yields better recovery and less invasive procedures. So it is now possible to provide more personnalized treatement to patient based on the very arrangement of their organs. A good radiotherapy plan must ensure that the target zone is properly irradiated while the organs at risk remain virtually untouched.

So far the existing method to derive this treatment plan are either very slow physics based model or rely on many approximation leading to a trial and error procedure degrading the final dose prediction capacity.

Our objective here is to assess the performances of deep learning based methods on such tasks. Given a CT scan of a patient, a region to irradiate and masks of their organs we want to predict the radiation dose. We will use the Mean Absolute Error (MAE) to assess our model.

2. Method

2.1. Architechture

Since the objective of the challenge is to predict the radiation dose value at each location of the authorized area, we needed an autoencoder model, whose output would be a single channel image. We have chosen to work with a UNet (Ronneberger et al., 2015) architecture, as you can see in the Figure 1. The UNet models, initially used in semantic segmentation, remain autoencoders and are adaptable to image regression. Moreover, the UNet architecture is widely used and has proven itself in image denoising (Gurrola-Ramos et al., 2021), thus showing that it is a good image synthesizer thanks to its skip connections. They allow to easily transport features from the source domain to the target domain by limiting information loss caused by the bottleneck.

We were interested in other autoencoder architectures such as Vision Transformers (Dosovitskiy et al., 2020) since these models have very competitive performances on vision tasks. However, they are known to be expensive to train, both in terms of computational capacity and data. We did not bother to look at regular autoencoders without skip connections as UNet have been shown to be consistently better.

Since our data set was composed of the CT, the mask of possible radiation and 10 organ masks for each case, we adapted the number of channels at the input of the model so that it can takes all the 12 images as one input. As for the output of the model, we have obviously chosen a single channel, corresponding to the prediction of the radiation dose.

For the training of the model we first used the loss MAE comparing the direct output of the model with the ground truth, without applying the mask of the possible radiation zone. We also tested a training where the MAE compared the model output with the applied mask and the ground truth.

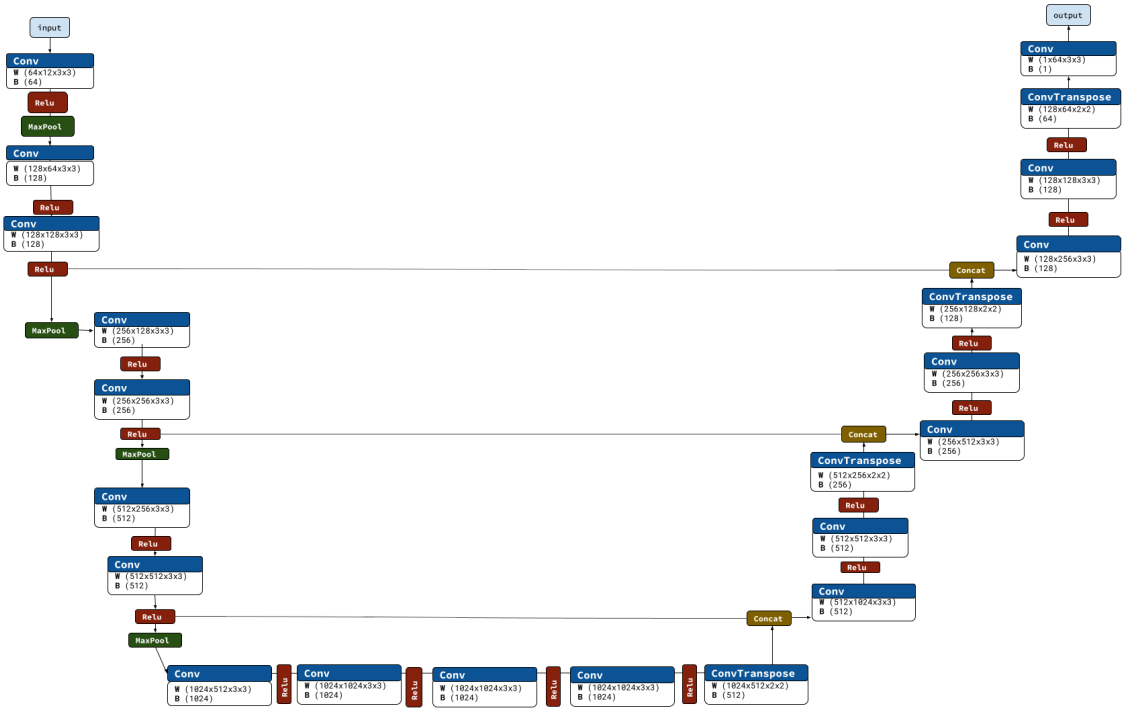

Figure 1: Our model

2.2. Dataset

In order to make our model generalizable and robust to different possible distributions of test data, we performed data augmentation on the training data. For this we applied geometric transformations such as random rotations and flips aswell as a random gaussian blur on the CT scan.

Also, in order to be generalizable to as many different input data as possible, and hopefully ensure that the result are not dependent of the acquisition procedure, we normalized all datasets according to the mean and standard deviation of the training data.

Since the set of CT images we had at our disposal could correspond to different organs of the human body, it seemed important to us to study the distribution of organs represented among the images in our dataset. Indeed, if some organs were poorly represented, it would have been difficult to correctly train a balanced model. The results of our our analysis is in Table 1. As you can see the distribution is uneven and some organs are underrepresented and it is divided the same way on all three datasets. Therefore it is not possible to rebalance between the validation and the training datasets. Based on these results, we have subsequently been careful to the performance of our model for each organ.

Organ number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Train	0.18	0.72	0.33	0.32	0.10	0.09	0.30	0.73	0.31	0.64
Validation	0.17	0.75	0.33	0.34	0.13	0.11	0.28	0.78	0.25	0.66
Test	0.21	0.76	0.32	0.31	0.07	0.10	0.37	0.75	0.35	0.68

Table 1: Organ distribution in the different datasets

3. Model tuning and comparison

3.1. First results

We trained all of our models with a Stochastic Gradient Descent optimizer, with a learning rate that increases linearly with the first 10 epochs and then decreases in a cosine decay following the workflow of Figure 2. After a first model trained on 100 epochs, we obtained a MAE score of 0.336046 on the test set. Note that we don't have access to the within submission standard deviation of the error on the test test ; on the validation set the MAE is of 0.392195 ± 0.338914.

In order to improve our model, we focused on the validation images with high MAE to determine the cases where our model was not performing well. We have displayed some of them and as you can see in Figure 3, our predictions are smoother than the ground truth because the model does not take into account the fact that the area is irradiated by the superposition of rays from different directions. One idea to remedy this was to impose a constraint on the gradient of our prediction so that it would be less smooth. We used Sobel filters to compute the gradient. If we denote GT the ground truth and P the prediction, the loss becomes :

$$\mathcal{L}(P, GT) = \text{MAE}(P, GT) + \lambda ||\nabla(P)||_2^2 \tag{1}$$

Unfortunately, this approach proved not to be efficient whilst adding a new hyper-parameter to tune.

After the study on the distribution of organs in our datasets, we verified that the performances were well balanced within the set of organs. As you can see in the Table 2 our model did not favor one type of organ even though some were underrepresented in the training dataset.

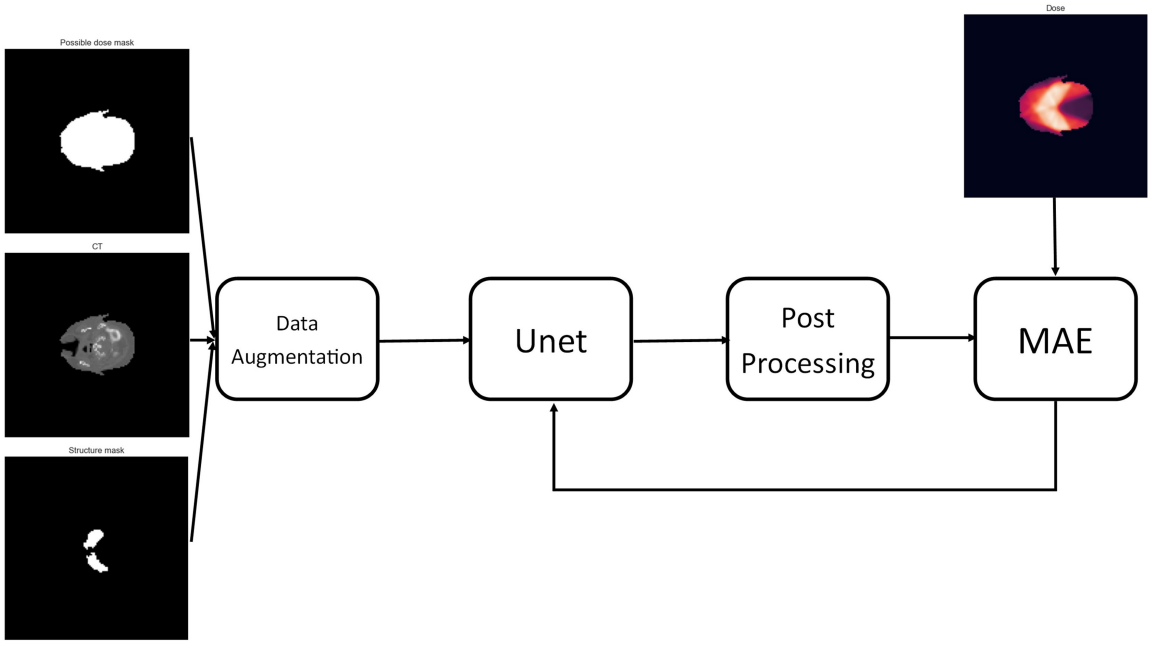

Figure 2: Workflow of our method. The steps of augmentation and post processing are optional per the procedure detailed in Table 3

Organ number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Validation	0.31	0.31	0.3	0.31	0.27	0.28	0.29	0.31	0.33	0.3

Table 2: MAE by organ

3.2. Ablation study

In the process of tuning our approach we experimented with different methods. The baseline is a Unet trained on 50 epochs with a base learning rate of 0.2. This model has no post processing involved and no augmentations. Note that the post processing during training only amounts to setting to zero the value of all pixels outside the possible dose mask. It is always done nonetheless during evaluation. The idea is that by doing it during training it could help the model to learn quicker in the early phases.

Given the results at Table 3, we used the final method and trained it on 100 epochs yielding a MAE score on the test set of **0.332832**.

4. Conclusion

Deep learning appears to be a suited method for such tasks. Yet it comes with two downsides. First, the major one, it has no guarantee to consistently produce accurate or safe treatment plan. Then, more deep learning related, it appears that despite changes in the workflow, the variation in results are very small and not necessarily significant.

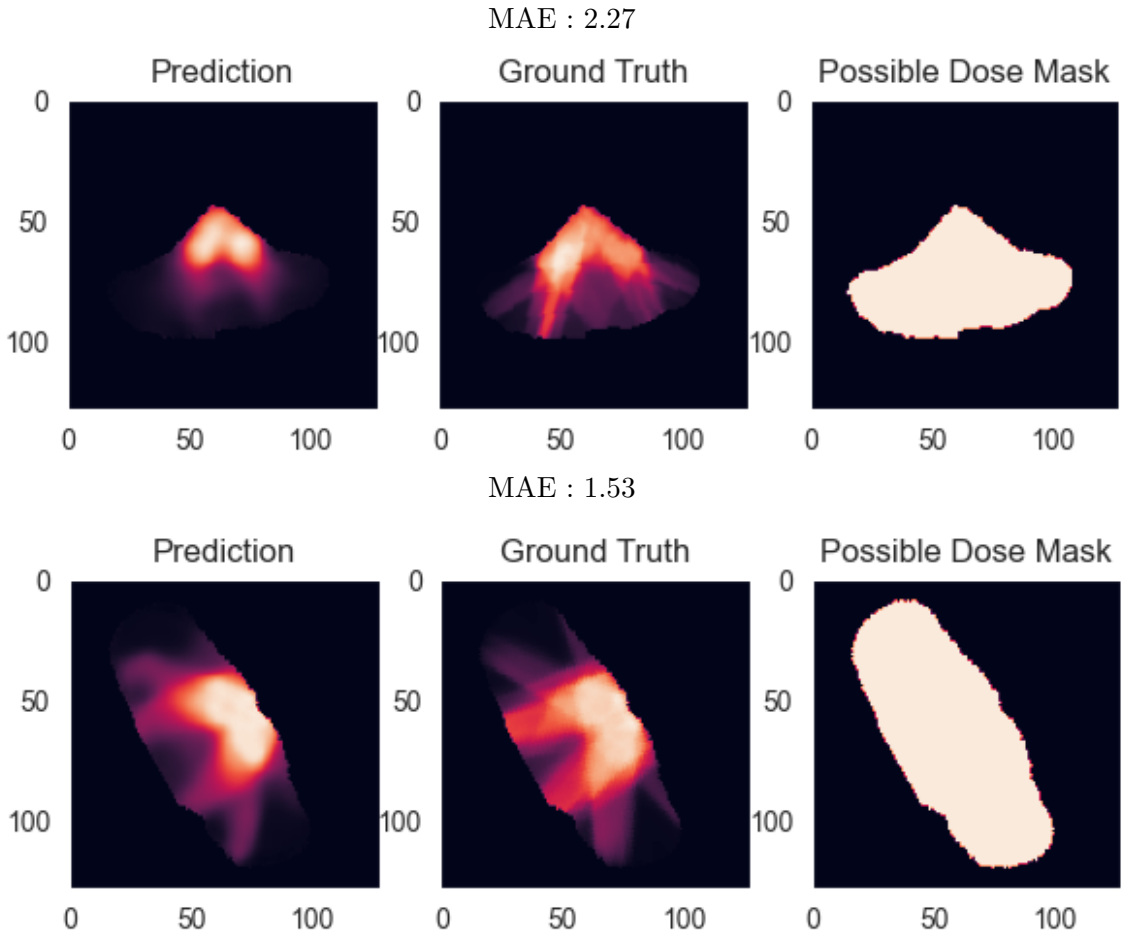

Figure 3: Example of images with high MAE

Methods	#Parameters	Aug	Post-process	Gradient pen	MAE
#1 (baseline)	49,8 M	×	×	×	0.355707
#2	49,8 M	✓	×	×	<u>0.348257</u>
#3	3,1M	✓	✓	×	0.365516
#4	49,8 M	✓	✓	✓	0.350957
Ours	49,8 M	✓	✓	×	**0.338081**

Table 3: Results in Mean Absolute Error scores of the ablation study. Bold denotes the best. result, underline indicates the best but one.

References

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Trans-

formers for image recognition at scale. *CoRR*, abs/2010.11929, 2020. URL https://arxiv.org/abs/2010.11929.

Javier Gurrola-Ramos, Oscar Dalmau, and Teresa E. M. Alarcón. A residual dense u-net neural network for image denoising. *IEEE Access*, 9:31742–31754, 2021.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. URL http://arxiv.org/abs/1505.04597.