Differentiable gamma index-based loss functions: accelerating Monte-Carlo radiotherapy dose simulation

Sonia Martinot[1,2,3] , Nikos Komodakis[4], Maria Vakalopoulou[2], Norbert Bus[1], Charlotte Robert[3], Eric Deutsch[3], Nikos Paragios[1]

[1]Therapanacea [2]CentraleSupelec [3]Institut Gustave Roussy [4]University of Crete

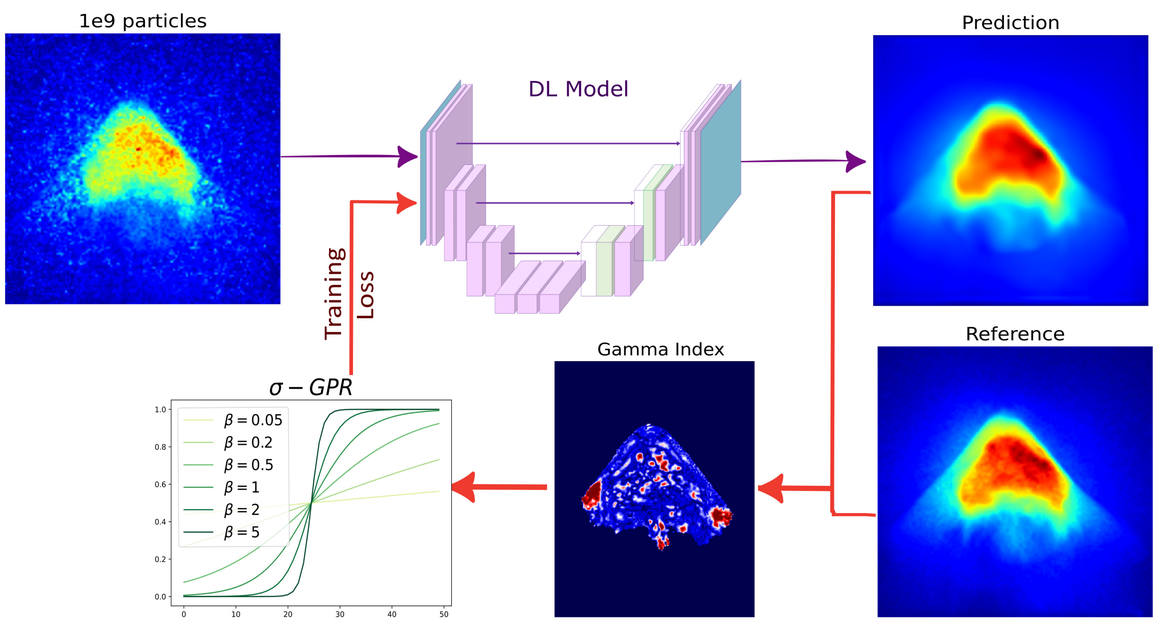
Introduction

- Monte-Carlo (MC) radiotherapy dose generation: **too slow** for clinical adoption.
- Possible Deep Learning (DL) **acceleration** of MC radiotherapy dose generation.
- DL models evaluated with clinical quality metric **Gamma index Passing Rate (GPR)**[1].
- **But GPR not suitable as loss function** for DL training because not differentiable.

Hypothesis: Training with *GPR-based loss functions yields more robust* DL models by circumventing solving a proxy problem.

Contributions

- Designed a **new class of GPR-based loss functions** overcoming non-differentiability of GPR.
- Proved hypothesis on the task of MC dose distribution denoising and benchmarked against other common loss functions.
- Brought 2D and 3D Gamma index **computations down to milliseconds** to enable fast DL training.



Definitions

Definition - Gamma index.

Let D_r and D_e be respectively the reference and the evaluated dose distributions. Each point in these distributions, P_r of D_r and P_e of D_e, has a coordinate vector, respectively $\vec{d}(P_r)$ and $\vec{d}(P_e)$, and a dose value, $D_r(P_r)$ and $D_e(P_e)$. Considering a point P_r in D_r with the points P_e in a vicinity $V(P_r)$ delimited by the Distance-to-Agreement (DTA), and Δ the tolerance on the relative dose difference, the Gamma index Γ is defined such that:

$$\forall P_r \in D_r, \Gamma(P_r) = \min_{P_e \in V(P_r)} \sqrt{\frac{||\vec{d}(P_e) - \vec{d}(P_r)||^2}{DTA^2} + \frac{(D_e(P_e) - D_r(P_r))^2}{\Delta^2}}$$

Definition - Gamma index Passing Rate.

*Let δ be a dose threshold and consider a point P_r of the reference distribution such that $D_r(P_r) \geq \delta$. Then, given a DTA and dose tolerance Δ, the evaluated distribution matches the reference at P_r, if the **passing criterion is satisfied, i.e. if** $\Gamma(P_r) \leq 1$. Then the GPR is the percentage of passing voxels in the evaluated distribution D_e:*

$$GPR(D_r, D_e) = \frac{\sum_{P_r \in D_r} \mathbb{1}_{D_r \geq \delta}(P_r) \cdot \mathbb{1}_{\Gamma \leq 1}(P_r)}{\sum_{P_r \in D_r} \mathbb{1}_{D_r \geq \delta}(P_r)}$$

Minimization problem:

- **Goal:** Find D_e closest to D_r such that $GPR(D_r, D_e) = 100\%$.
- Maximizing the GPR amounts to minimizing $L^{\delta}_{GPR}(D_r, D_e) = 1 - GPR(D_r, D_e)$.
- L^{δ}_{GPR} has zero-gradients everywhere jeopardizing backpropagation.

Sigmoid-GPR loss function

Definition. *Let σ be the sigmoid function such that $\sigma(x) = \left(1 + \exp^{-\beta x}\right)^{-1}$ with sharpness parameter β. Then we define $L^{\delta}_{\sigma-GPR}$ as follows:*

$$L^{\delta}_{\sigma-GPR}(D_r, D_e) = 1 - \frac{\sum_{P_r \in D_r} \sigma(\beta \cdot (1 - \Gamma(P_r)))\mathbb{1}_{D_r \geq \delta}(P_r)}{\sum_{P_r \in D_r} \mathbb{1}_{D_r \geq \delta}(P_r)}$$

- $L^{\delta}_{\sigma-GPR} \to L^{\delta}_{GPR}$ as $\beta \to +\infty$.
- Annealing schedule of hyperparameter β to mitigate vanishing gradients during training.

- Gamma index computation time as fast as SSIM yields fast and tractable training durations.

Metric	3D dose Time(ms)	2D dose Time(ms)
SSIM	27.49 ± 7.73	5.06 ± 3.32
$L^{\delta}_{\sigma-GPR}$, $L^{\delta}_{\Gamma>1}$, L^{δ}_{Γ}	**30.54 ± 0.01**	**4.51 ± 0.00**
Exhaustive	985 ± 515	8.01 ± 2.28
PyMedPhys	> 1 second	> 100 ms

Table 1. Speed comparison of metrics computed over 2D or 3D dose distributions.

Experimental design

Goal: Denoise low quality MC dose (1e9 simulated particles) to match corresponding high quality MC dose (1e11 simulated particles).

- **Dataset** [2]: 50 patients with 3D dose distributions amounting to 11k 2D training samples.
- **Preprocessing**: Normalization using maximum dose value over training set. Padding up to size 256×256.
- **Data augmentation**: vertical and horizontal flipping.
- **Model**: 2D UNet-like architecture.
- **Annealing schedule of** β: starting with low value $\beta = 2 \times 10^{-2}$, slow increase over training up to $\beta_{max} = 5$.
- **Asymptotic comparison** by also training on L^{δ}_{Γ} and $L^{\delta}_{\Gamma>1}$ corresponding to small values of β.
- **Benchmark** against trainings with the SSIM, MAE, MSE, SSIM+MSE or SSIM+MAE.

Results

- Models trained on GPR-based functions **outperform other trainings** on MAE, MSE and GPR.
- Models trained on GPR-based functions are **more robust to outliers**.
- SSIM not suited to reflect quality of generated dose distribution.

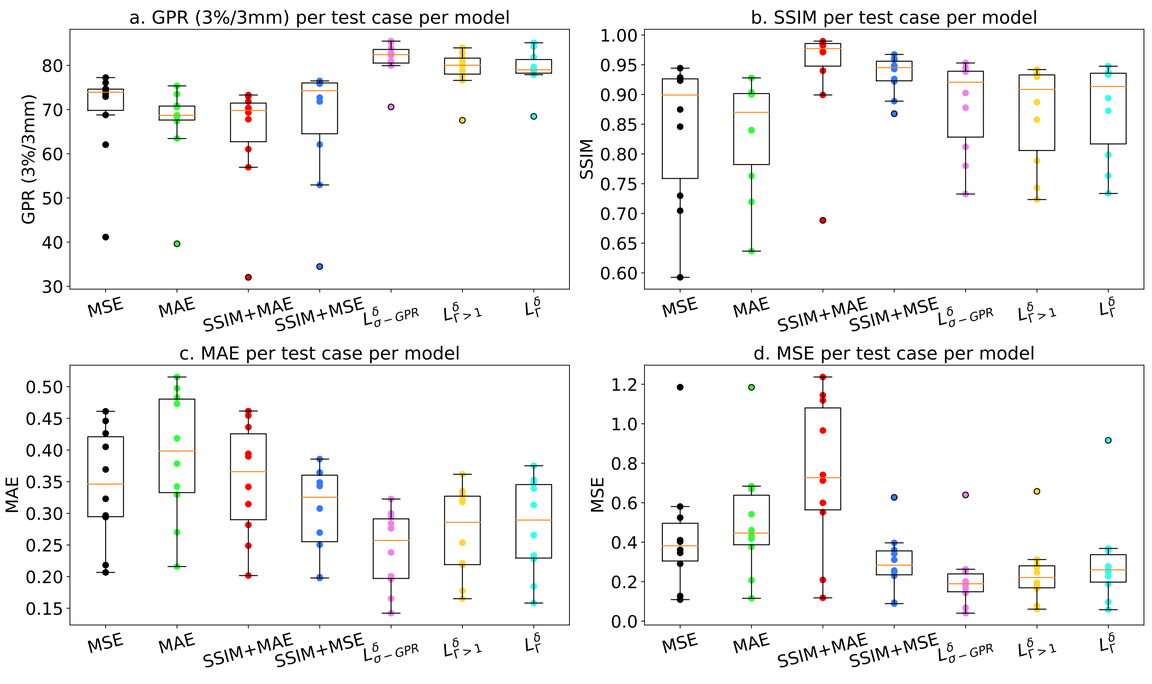


Figure 1. Boxplots of evaluation metrics afor each trained models on the test set depending on the loss function used. The x axis specifies the loss function with which the corresponding model was trained.

Loss function	GPR 2%/2mm	GPR 3%/2mm	GPR 3%/3mm	SSIM (%)	MAE	MSE
MAE	47.6 ± 7.6	53.8 ± 8.4	66.5 ± 10.0	83.3 ± 9.2	0.39 ± 0.10	0.51 ± 0.28
MSE	50.2 ± 7.8	56.7 ± 8.7	69.5 ± 10.3	83.9 ± 11.5	0.34 ± 0.09	0.43 ± 0.29
SSIM+MAE	46.3 ± 8.8	52.3 ± 9.8	64.6 ± 11.9	**94.0 ± 8.8**	0.35 ± 0.09	0.74 ± 0.36
SSIM+MSE	49.1 ± 10.0	55.3 ± 11.1	67.5 ± 13.2	93.3 ± 3.1	0.30 ± 0.07	0.30 ± 0.15
L^{δ}_{Γ}	57.7 ± 3.3	65.0 ± 3.5	79.2 ± 4.3	87.5 ± 7.6	0.28 ± 0.07	0.30 ± 0.23
$L^{\delta}_{\Gamma>1}$	57.2 ± 3.9	64.6 ± 4.1	79.0 ± 4.4	86.7 ± 8.1	0.27 ± 0.07	0.25 ± 0.16
$L^{\delta}_{\sigma-GPR}$	**59.3 ± 3.3**	**66.8 ± 3.6**	**81.4 ± 4.0**	88.2 ± 7.5	**0.24 ± 0.06**	**0.22 ± 0.16**

Table 2. Evaluation metrics over the test dose distributions depending on the loss function used fo training. With bold we indicate best performing methods per metric.

Key Takeaways

- **First GPR-based loss functions** intended for training both 2D and 3D DL models.
- Models trained with Sigmoid-GPR more robust and outperform models trained with other usual loss functions.
- **Future work:** improvement of implementation and application to other tasks.

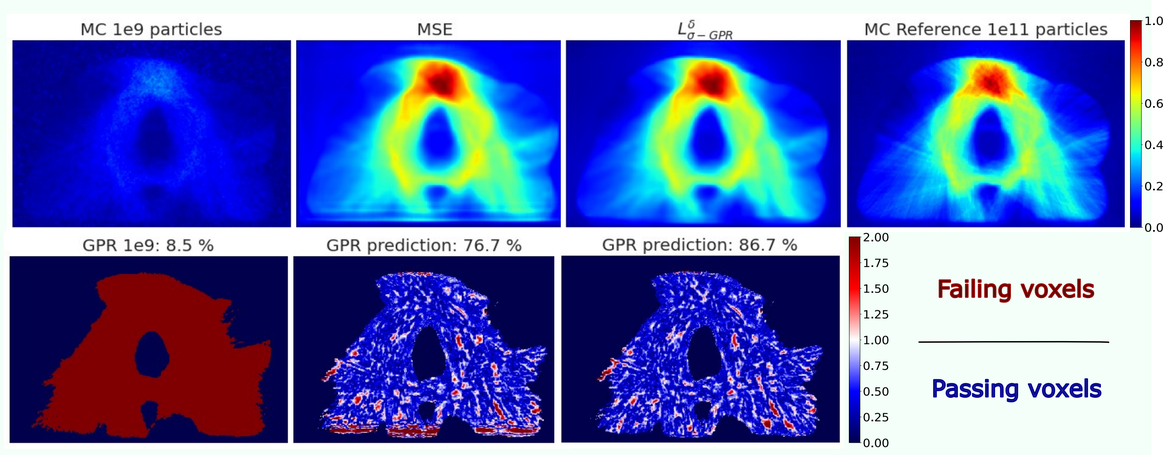


Figure 2. First row from left to right: slice of the 1e9 dose volume, predictions of models trained with MSE and $L^{\delta}_{\sigma-GPR}$, and reference 1e11 dose. Second row: corresponding gamma index maps.

References

[1] D. Low, W. Harms, S. Mutic, and J. Purdy, "A technique for the quantitative evaluation of dose distributions," *Medical physics*, vol. 25, no. 5, pp. 656–661, 1998.

[2] S. Martinot, N. Bus, M. Vakalopoulou, C. Robert, E. Deutsch, and N. Paragios, "High-particle simulation of monte-carlo dose distribution with 3d convlstms," in *MICCAI*, pp. 499–508, 2021.

sonia.martinot@hotmail.fr