Automated Treatment Planning in Radiation Therapy using Generative Adversarial Networks

Rafid Mahmood[1] Aaron Babier[1] Andrea McNiven[2] Adam Diamant[3] Timothy C. Y. Chan[1]

[1]Mechanical & Industrial Engineering, University of Toronto [2]Princess Margaret Cancer Centre [3]Schulich School of Business, York University

Mechanical & Industrial Engineering UNIVERSITY OF TORONTO

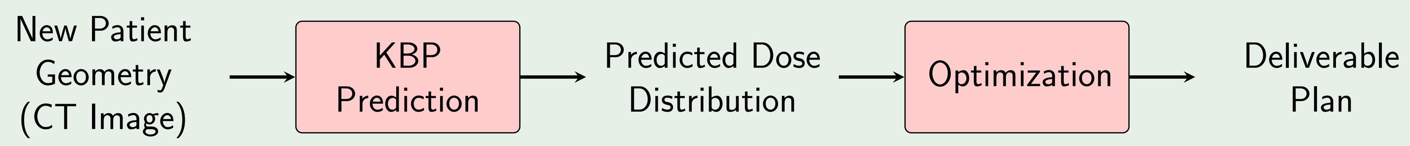
Knowledge-Based Treatment Planning in Radiation Therapy

Radiation therapy (RT) is used to treat over 50% of all cancer cases. The work-flow is:

1. A dosimetrist generates plans by solving an optimization problem using patient CT images.
2. An oncologist reviews the plan and suggests changes to the dosimetrist.

This process repeats until the oncologist approves the plan, which can take several days.

Knowledge-based planning (KBP) automates plan generation by predicting a dose distribution using historical cases and then correcting it to a "deliverable" plan via optimization.



Our Contributions

We train a generative adversarial network (GAN) to color CT images. Specifically, we:

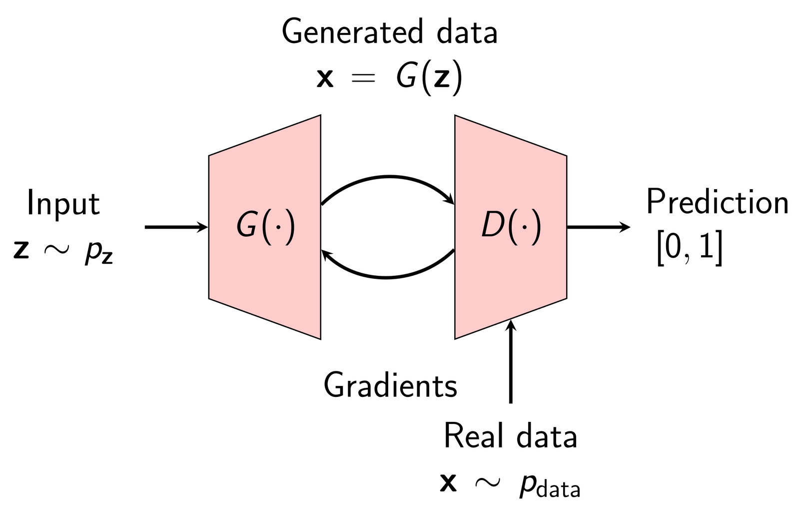
1. Develop the first GAN-based KBP prediction model and show that it outperforms state-of-the-art benchmarks (e.g., ML techniques, deep neural networks) in several clinical metrics.
2. Demonstrate on a oropharyngeal cancer data set that our GAN-based KBP can successfully generate plans that often meet prescription criteria more than clinical treatment plans.

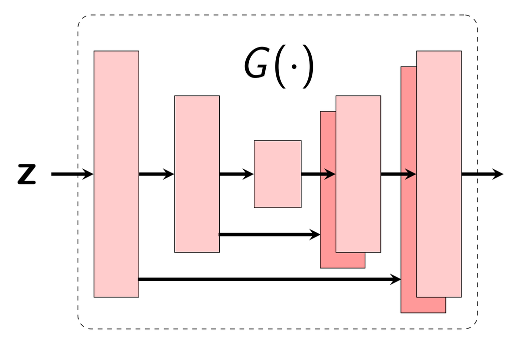
Generative Adversarial Networks (GANs)

A generator $G(\cdot)$ takes $\mathbf{z} \sim p_{\mathbf{z}}$ and outputs a sample $\mathbf{x} = G(\mathbf{z})$. A discriminator $D(\cdot)$ takes a real and generated sample $\mathbf{x}$ and predicts which is real ($D(\mathbf{x}) = 1$) and which is fake ($D(\mathbf{x}) = 0$).

Style Transfer GANs map an image in style $\mathbf{z}$ to a different style $\mathbf{x}$, e.g., grayscale to color. They are trained by minimizing reconstruction loss that penalizes the distance from ground truth.

$$\min_G \max_D \left\{ \underbrace{\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}}[\log(1 - D(G(\mathbf{z})))]}_{\text{GAN loss}} + \lambda \underbrace{\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}, \mathbf{z}\sim p_{\mathbf{z}}}[\|\mathbf{x} - G(\mathbf{z})\|_1]}_{\text{reconstruction loss}} \right\}.$$

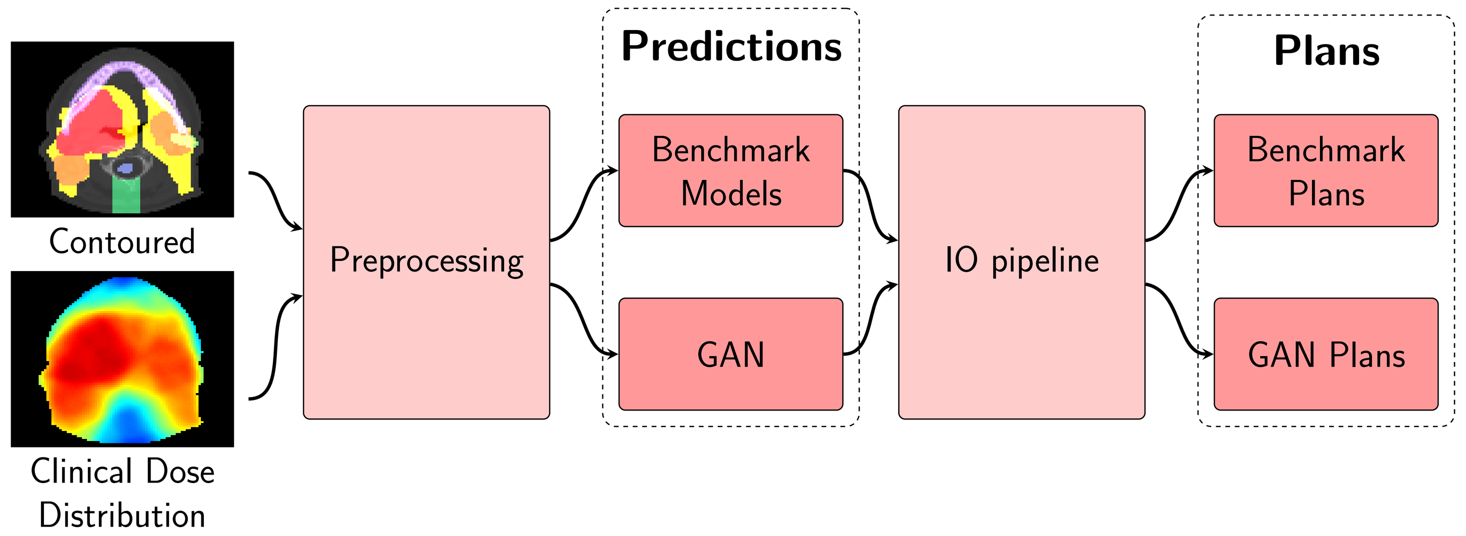




Style Transfer GAN Generator is a U-net (encoder-decoder with skip connections)

Dataset

CT scans and historical treatment plans for 217 patients with oropharyngeal cancer. Plans were prescribed 70 Gy, 63 Gy, and 56 Gy to high, intermediate, and low risk targets, respectively.



Method for Generating Treatment Plans

Preprocessing: 3D patient volumes were divided into voxels ($4mm \times 4mm \times 2mm$); each voxel was labeled as organ-at-risk (OAR), planning target volume (PTV), or unclassified.

Predictions: 3D CT images were split into 2D slices ($128 \times 128px$). We used 130 plans (15,675 images) to train the GAN and benchmark models. We predicted plans for 87 out-of-sample patients.

IO Pipeline: An optimization model converted predicted dose distributions into "deliverable" plans.

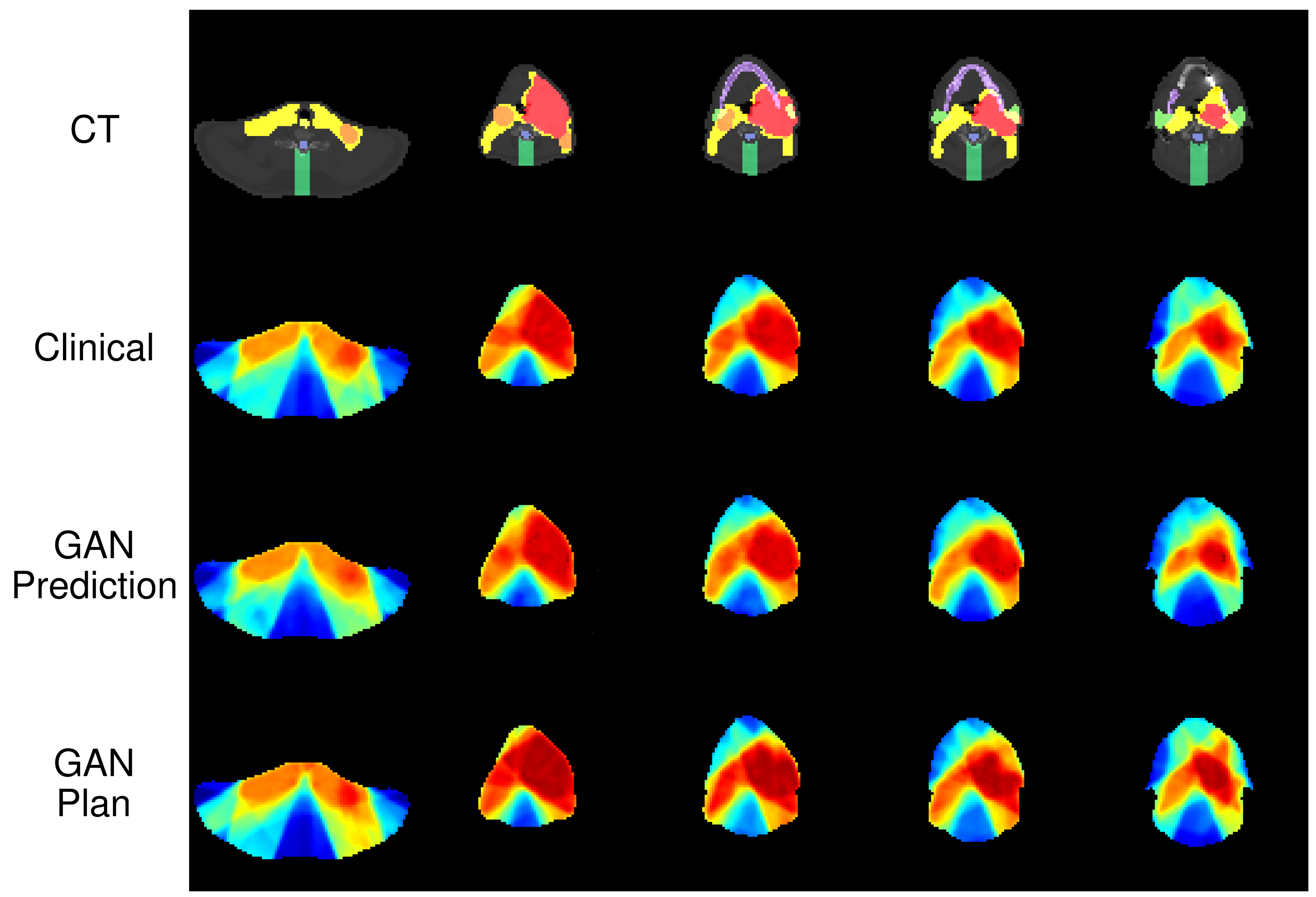
Benchmark KBP Approaches

Bagging Query (BQ): A look-up table identifies deliverable plans of patients with similar geometries and uses low-dimensional summaries of their 3D dose distributions to the targets and OARs.

Generalized PCA (gPCA): Combination of PCA and linear regression using patient geometry features to predict low-dimensional dose distribution summaries for targets and OARs.

Random Forest (RF): Predicts dose to each voxel using custom patient geometry-based features.

U-net Convolutional Neural Network (CNN): Predicts dose to each voxel in 2D slices from CT images using a U-net convolutional neural network architecture.



Results and Clinical Criteria

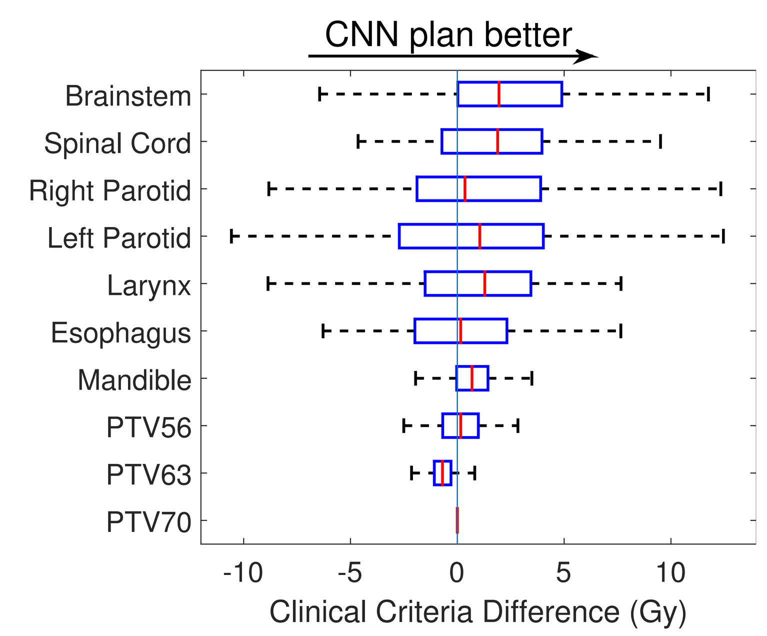
All methods were evaluated on the percentage of the final plans that satisfied 10 clinical criteria (seven for OARs and three for PTV).

OAR Sparing Criteria: One criteria per OAR restricts either the mean or the max dose to the region to be below a certain threshold value, e.g., max dose to any voxel in the brainstem should be less than 54 Gy.

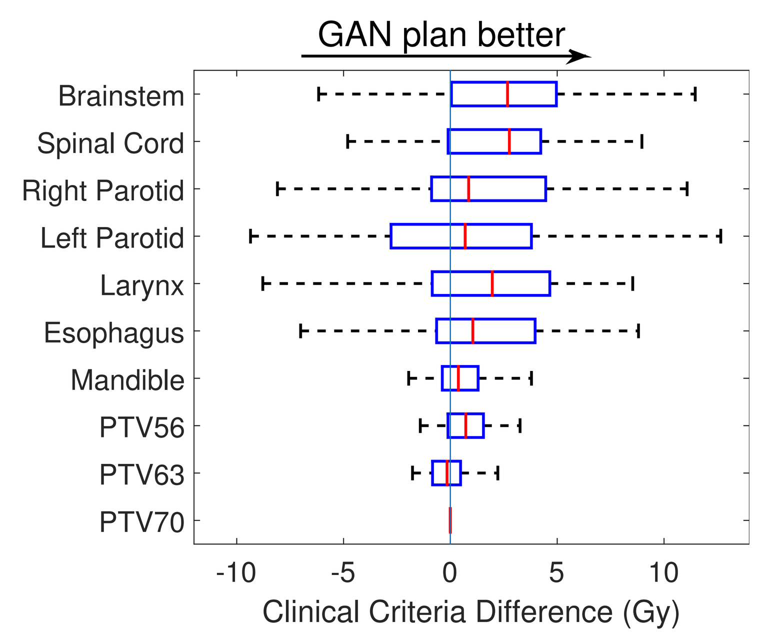
Target Coverage Criteria: One criteria per target ensures it receives the prescribed dose, e.g., 99% of all voxels in the PTV70 should receive 70 Gy.

	BQ	gPCA	RF	CNN	GAN	Clinical
OAR Criteria	61.6%	65.8%	71.5%	72.5%	72.8%	72.0%
PTV Criteria	83.5%	85.7%	68.0%	76.3%	81.3%	76.8%
All Criteria	67.6%	71.2%	70.7%	73.6%	75.2%	73.3%

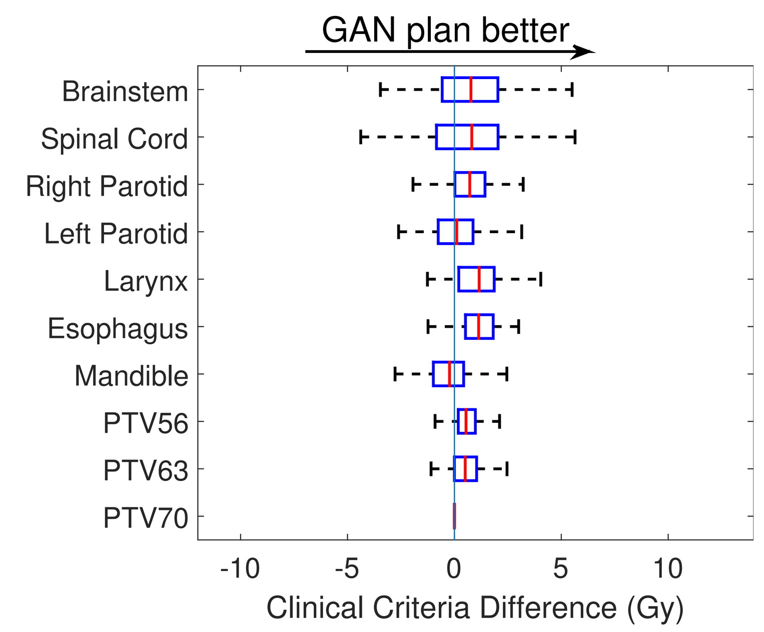
The GAN plans were best at sparing OARs, delivered the requisite amount of radiation to the target, and were most likely to satisfy all clinical criteria.



(a) CNN − clinical



(b) GAN − clinical



(c) GAN − CNN

Head-to-head comparisons of plans. The x-axis represents difference in dose (measured in Gy), normalized for criteria condition. Box boundaries represent 75th and 25th percentiles.

Conclusions

We demonstrate that our GAN-based KBP work-flow:

1. Predicts 3D dose distributions without cancer-site specific feature engineering.
2. Outperforms all benchmark approaches for one of the hardest cancer treatments to plan; we expect these results will hold for simpler more common cancers.

Email: rmahmood@mie.utoronto.ca