OpenKBP Grand Challenge

February 21, 2020

1 Background

The Open Knowledge-Based Planning (OpenKBP) Grand Challenge provides an augmented variation of real clinical data. Specifically, we provide patient data (e.g., CT images) from several institutions that is available on The Cancer Imaging Archive (TCIA) Clark et al. [2013], which hosts several open-source datasets. The data is augmented with dose distributions from *synthetic* plans that are generated via a variation of a published automated planning method Babier et al. [2020]. This document outlines the contents of the dataset and how to use it.

2 About the data

The data we provide is structured in a way to facilitate the development and validation of dose prediction models. Specifically, we cleaned the data from TCIA and extracted a set of CT images and structure masks (i.e., the voxels included in each structure). Next, the image data for each patient was downsampled to $128 \times 128 \times 128$ voxel tensor. Each patient has slightly different voxel dimensions (around 3.5 mm $\times$ 3.5 mm $\times$ 2 mm), but the exact dimensions of every voxel in a patient are also provided.

2.1 Organs-at-risk

We used the original clinical contours to segment organs-at-risk (OARs) and create their appropriate masks. The labeled OARs in this dataset are the brainstem, spinal cord, right parotid, left parotid, larynx, esophagus, and mandible. Masks were omitted if the OAR was not contoured in the original clinical plan.

2.2 Targets

Plans were prescribed 70 Gy, 63 Gy, and 56 Gy in 35 fractions to the gross disease (PTV70), intermediate-risk target volumes (PTV63), and elective target volumes (PTV56). Every PTV was clipped to be no closer than 5 mm from the surface of the patient. For the purpose of modelling, the provided PTV masks have no overlap with each other. Every plan has either (i) a PTV70; (ii) a PTV56 and PTV70, or (iii) a PTV56, PTV63, and PTV70.

2.3 Dose generation

The provided dose distributions are from fluence-based treatments plans with similar degrees of fluence complexity Craft et al. [2007]. These fluence plans are "delivered" by a collection of *beamlets* that deliver the dose to the patient based on a *dose deposition matrix* (i.e., plan physics). To ensure consistent physics, all dose deposition matrices were made using the same parameters in a Computational Environment for Radiotherapy Research (CERR) Deasy et al. [2003]. Also, all plans were also delivered from nine equispaced coplanar fields at 0°, 40°, ..., 320° with 6 MV step-and-shoot intensity-modulated radiation therapy in 35 fractions to satisfy the prescribed dose to the high-risk target (i.e., PTV70).

3 Summary of database

The data for each patient is encoded as comma-separated values (CSV) files, which are separated into directories with the corresponding patient number. The files for each patient include:

dose.csv: the full 3D dose distribution that was used to treat the patient (in units of Gy)

ct.csv: grey-scale images of the patient prior to treatment.

voxels.csv: The size of the patient voxel (in units of mm)

possible_dose_mask.csv: a mask of voxels that can receive dose (i.e., the dose will always be zero where this mask is zero).

Structure masks: a boolean tensor that labels any voxel that is contained in the respective structure. The tensor for each structure is stored as a CSV file named after the appropriate structure. Note that only structures that were contoured in the patient have CSV files.

Brainstem.csv: mask of brainstem voxels

SpinalCord.csv: mask of spinal cord voxels

RightParotid.csv: mask of right parotid voxels

LeftParotid.csv: mask of left parotid voxels

Esophagus.csv: mask of esophagus voxels

Larynx.csv: mask of larynx voxels

Mandible.csv: mask of mandible voxels

PTV56.csv: A target that should receive 56 Gy

PTV63.csv: A target that should receive 63 Gy

PTV70.csv: A target that should receive 70 Gy

For the most part, this data is standard in most dose prediction models. The exception is the "possible_dose_mask.csv", which identifies where the dose can be nonzero based on our *dose deposition matrix* calculated by CERR. To use this feature, simply force any dose that is outside of this mask to zero. Doing this will also bound the file size of submissions and ensure that submissions can be made to CodaLab without exceeding the submission file size limit, which was the primary motivation for including the "possible_dose_mask.csv".

3.1 Loading the data

We strongly recommend using the data-loader provided in our GitHub repository. The data loader will load the dose, CT, and masks as 3D tensors that are 128x128x128 voxels. The rest of this section outlines the format of the data for participants interested in building their own data loader or using a coding language other than Python. Please email any questions to openkbp@gmail.com.

3.2 Data format

The data formatted as sparse matrices; only data points with non-zero values are provided. The first column of each CSV file is a list of indices, and the second column, which is also labeled data, is a list a values for their corresponding indices. Those indices are stored as single numbers that map to a 3D (i.e., x-y-z) coordinate system. Each index is a single number that can be unraveled into an x, y, and z coordinate in one of two ways:

Python users: Use the unravel_index function from numpy with the dims parameter equal to the shape of the patient (128, 128, 128) and the order parameter equal to 'C'. Similar functions should be available in other languages. The ravel_multi_index may be used to convert those x, y, and z coordinates back into a vector of indices.

Non-Python users: the unraveled_indices.csv contains the x, y, and z coordinates for any index between 1 and 128^3 at its respective row.

4 Submitting results

Results must be submitted as a zip file (with any name) to Coda lab. We strongly recommend zipping the directories with a python function, which we provide in the main script and notebook, because some programs add extra hidden files that will crash the online evaluation stage. Within that directory, information for each patient must be saved as a CSV file named after the patient it is for (e.g., *pt_1.csv*). This means that a submission to the validation phase should be a zipped directory with about 30 CSV files named after the patients in the validation set. The CSV can take one of two formats, but must be consistent:

3D dose submissions: Must be the same format as the dose distributions that are provided. The first column indexes any non-zero values and the second column, with the data heading, lists the corresponding non-zero values.

DVH submissions: The DVHs must be converted into a table where each row lists the value for a metric. Specifically, the first column is the name of the metric-roi pair (e.g., *('mean', 'RightParotid')* is the string that maps to the mean dose to the right parotid) and the second column is the corresponding value (e.g., *30*, which would imply a prediction of 30 Gy). This is identical to the format of saving a row of the pandas data frame that is returned by the calculate_metrics method within the EvaluateDose class. An example CSV file is also included on CodaLab.

5 Evaluation

Submissions are evaluated with two scores: (1) the *Dose score* and (2) the *DVH score*. Submissions to the 3D dose stream will be scored on both metrics automatically, but submission to the DVH stream will only be scored on the DVH score since calculating a Dose score from a DVH is impossible. The scoring metrics are summarized below:

5.1 Dose score

The dose score is meant to provide a general measure of prediction quality. It measures the mean absolute difference between predictions and the ground truth dose distributions. The evaluation is only completed over voxels within the "possible_dose_mask".

5.2 DVH score

The DVH score is meant to provide a clinical measure of prediction quality. It measures the mean absolute difference between the predictions and the ground truth dose distributions only at important DVH metrics. The following aspects of the DVH curve are evaluated for each OAR:

$\mathcal{D}_{0.1cc}$**:** The maximum dose to 0.1cc (or 100mm^3)

$\mathcal{D}_{mean}$**:** The mean dose (i.e., area under DVH curve)

The DVH for each target is evaluated over a different set of metrics:

$\mathcal{D}_1$**:** The minimum dose received by all but 0.1cc of the target (or 100mm^3)

$\mathcal{D}_{95}$**:** The minimum dose received by 95% of the target (i.e., 5^{th} percentile)

$\mathcal{D}_{99}$**:** The minimum dose received by 99% of the target (i.e., 1^{st} percentile) These metrics are only evaluated over structures that exist in a patient. If no structure exists the difference is recorded as a blank space or NaN.

References

Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, Lawrence Tarbox, and Fred Prior. The cancer imaging archive (tcia): maintaining and operating a public information repository. *J Digit Imaging*, 26(6):1045–57, Dec 2013. doi: 10.1007/s10278-013-9622-7.

Aaron Babier, Rafid Mahmood, Andrea L McNiven, Adam Diamant, and Timothy C Y Chan. Knowledge-based automated planning with three-dimensional generative adversarial networks. *Med Phys*, 47(2):297–306, Feb 2020. doi: 10.1002/mp.13896.

D. Craft, P. Suss, and T. Bortfeld. The tradeoff between treatment plan quality and required number of monitor units in intensity-modulated radiotherapy. *Int J Radiat Oncol Biol Phys*, 67 (5):1596–605, 2007.

J. O. Deasy, A. I. Blanco, and V. H. Clark. CERR: a computational environment for radiotherapy research. *Med Phys*, 30(5):979–85, 2003.