initial data
DICOM files

cohort folder

patient_1

patient_2

patient_3

1 standardize architecture

Patient_id

Date

Serie number

2 convert dicom to nifti

CT

RTSTRUCT

→

image.nii.gz

mask_1.nii.gz

mask_2.nii.gz

mask_3.nii.gz

check voxel intensities

3 tidy files by type

serie folder

DICOM

NIFTI

4 missing data

- CT serie without RTSTRUCT
- missing nifti present in rtstruct
- tumor mask missing ring mask
- mask baseline missing mask evaluation
- dupplicates : same patient id, same date, same location, same ROI num, same ROI type (tum/ring)
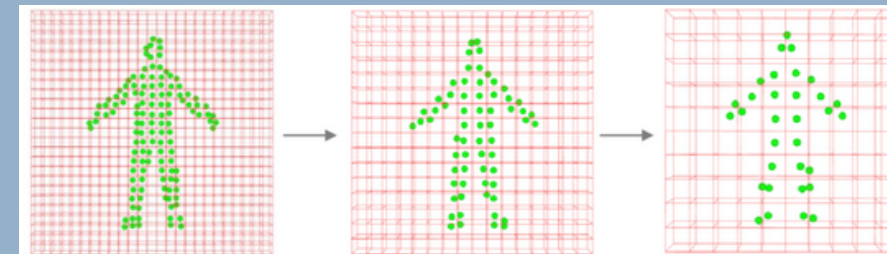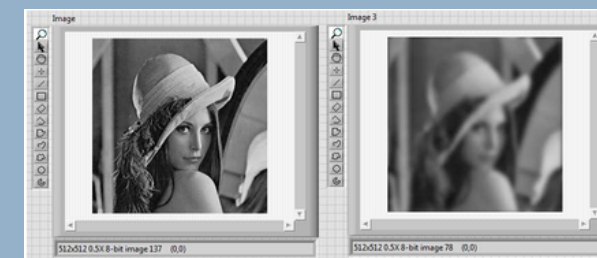
Detect problems in you data

5 preprocess nifti

- resample voxel size (mandatory)
- rescale voxel intensity
- apply spatial filter
- discretization of voxel intensity
- delete nifti not resampled

Standardize images and masks across series and patients

https://arxiv.org/html/2402.17521

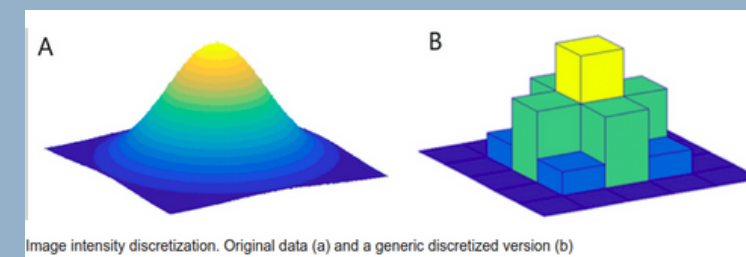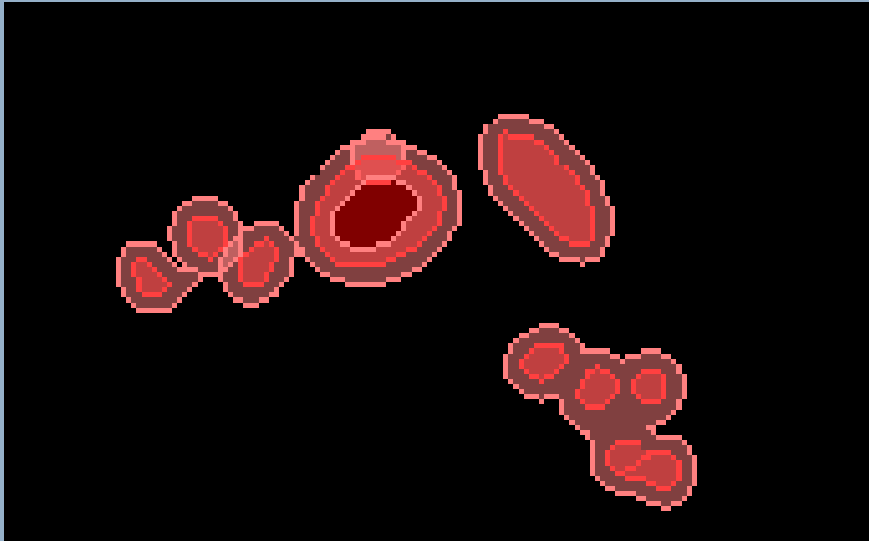https://forums.ni.com/t5/Machine-Vision/gaussian-filter

Image intensity discretization. Original data (a) and a generic discretized version (b)

https://www.researchgate.net/figure/mage-intensity-discretization-Original-data-a-and-a-generic-discretized-version-b_fig1_343613289

delete external contours
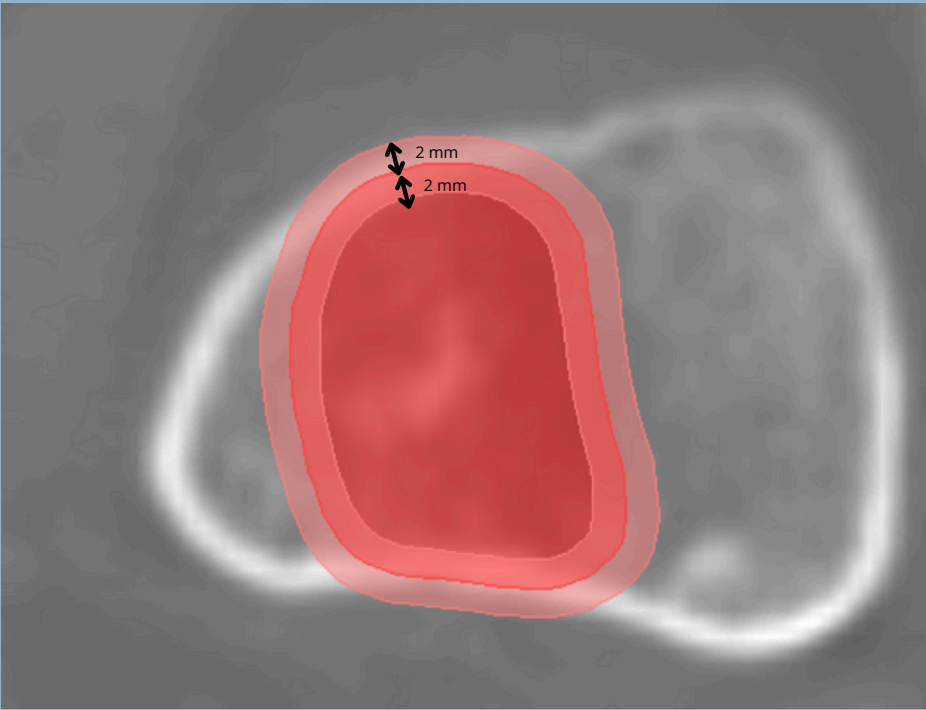
delete lesions too small

6 check mask components

decompose tumor masks with several components into a single mask/file per component

delete ring mask with no corresponding ring mask

7 check ring

9 check and correct ring

Create ring when ring incorrect and replace original file

8 create ring