

Python for Data Science: NumPy und SimpleITK

Bei allen Aufgaben: Fragen Sie mich, wenn Sie etwas nicht verstehen oder nicht weiterkommen!

Im Rahmen dieser Übung werden Sie einen 3D-Medizinbild-Datensatz untersuchen und in mehreren Schritten verarbeiten. Der Datensatz ist ein Thoraxbild, das durch eine CT-Untersuchung erzeugt wurde.

Beantworten Sie die Fragen in der Aufgabenstellungen direkt im Code durch eine entsprechende Ausgabe, z.B.:

```
print("Das Ergebnis der Berechnung 2*3 ist", 2*3)
```

Anleitung:

- Auf der Moodle-Seite dieses Aufgabenblatts finden Sie die Bilddaten.
- Schauen Sie sich das Bild mit MITK an (installiert auf den Poolrechnern, s.a. www.mitk.org).
- Für diese Aufgabe benötigen wir das Python Paket SimpleITK (s.a. simpleitk.org). SimpleITK können Sie – nachdem Sie die conda-Umgebung aktiviert haben – mit folgendem Befehl installieren:

```
pip install SimpleITK
```
- Üblicherweise wird SimpleITK in Python als `sitk` importiert.
- Die SimpleITK-Funktion zum Einlesen von Bildern heißt `ReadImage`. Sie liefert ein `SimpleITK.Image`-Objekt zurück.
- Mit der Methode `GetSpacing()` des Image-Objekts können Sie sich die Größe der Voxel in x-, y- und z-Richtung in Millimetern zurückgeben lassen.
 - Sind die Voxel Würfel oder Quader mit unterschiedlichen Seitenlängen?
- Mit der Methode `GetSize()` des Image-Objekts können Sie sich die Anzahl der Voxel in x-, y- und z-Richtung zurückgeben lassen.
 - Wie viele Voxel hat das Bild in den drei Raumrichtungen?
 - Welchen Datentyp gibt die Methode `GetSize()` zurück?
- Lassen Sie die Antworten auf die folgenden Fragen berechnen OHNE dafür selbst eine Schleife zu programmieren (und auch OHNE selbst mehrfach auf einzelne Elemente zuzugreifen):
 - Wie groß ist der Bildausschnitt in die drei Raumrichtungen *in Millimetern*?
- Die Lungen einschließlich der Luftwege können bei diesem Bild recht gut mit der Funktion `sitk.ConnectedThreshold(image, seedList, lower, upper)` extrahiert (segmentiert) werden: Dabei werden alle Voxel ab einem (oder mehreren) Startpunkten (sog. seeds) gesucht, mit den Startpunkten über einen Weg verbunden („connected“) sind, deren Helligkeitswerte zwischen einem angegebenen unteren (`lower`) und oberen (`upper`) Wert liegen. Die gefundenen Voxel werden dabei auf 1 gesetzt, alle anderen auf 0. Verwenden Sie als Startpunkt den Punkt (120,140,100) (`seedList` muss allerdings eine Liste von Punkten sein) und `lower=-1100` und `upper=-200`. Speichern Sie das Ergebnis als `lungs.nrrd` und schauen Sie es sich in MITK an.
- Konvertieren Sie das Ergebnis mit `sitk.GetArrayFromImage(image)` in ein NumPy-Array. Benutzen Sie dieses NumPy-Array um die Anzahl der Voxel der Lunge (einschließlich der Luftwege) zu bestimmen. Rechnen Sie die Anzahl in Liter um. Das Lungenvolumen eines erwachsenen Menschen beträgt durchschnittlich 5 bis 6 Liter.

- Lassen Sie sich die Häufigkeit der Helligkeitswerte, das sog. Histogramm, des CT-Bildes anzeigen. Verwenden Sie dazu die Funktion `hist(x, bins)` von `matplotlib`. Diese braucht als Eingabe eine eindimensionale Liste `x` von Werten (kann also nicht direkt mit Bildern arbeiten). Ein mehrdimensionales NumPy-Array kann mit der Methode `flatten()` in eindimensionale NumPy-Array konvertiert werden. Verwenden Sie `bins=100`.
- Beschriften Sie die Achsen.
- Lassen Sie sich das Histogramm der zur Lunge (einschließlich der Luftwege) gehörenden Werte des CT-Bildes anzeigen – welche Helligkeitswerte treten also innerhalb der Lunge (einschließlich der Luftwege) auf? Verwenden Sie wieder `bins=100` und beschriften Sie wieder die Achsen.

Tipp: Die oben erstellte Segmentierung enthält ja innerhalb der Lunge/der Luftwege den Wert 1, sonst 0. Sie müssen mehr tun, als nur `hist` aufrufen.