Histograms should not be used to segment functional lung MRI

Nicholas J. Tustison[1], Talissa A. Altes[2], Kun Qing[3], James C. Gee[4], G. Wilson Miller[1], John P. Mugler III[1], Jaime F. Mata[1]

[1]Department of Radiology and Medical Imaging, University of Virginia, Charlottesville, VA
[2]Department of Radiology, University of Missouri, Columbia, MO
[3]Department of Radiation Oncology, City of Hope, Duarte, CA
[4]Department of Radiology, University of Pennsylvania, Philadelphia, PA

Corresponding author:
Nicholas J. Tustison, DSc
Department of Radiology and Medical Imaging
University of Virginia
ntustison@virginia.edu

Abstract

Magnetic resonance imaging using hyperpolarized gases has facilitated the novel visualization of airspaces, such as the human lung. The advent and refinement of these imaging techniques have furthered research avenues with respect to the growth, development, and pathologies of the pulmonary system. In conjunction with the improvements associated with image acquisition, multiple image analysis strategies have been proposed and developed for the quantification of hyperpolarized gas images with much research effort devoted to semantic segmentation, or voxelwise classification, into clinically-oriented categories based on functional ventilation levels. Given the functional nature of these images and the consequent complexity of the segmentation task, many of these algorithmic approaches reduce the complex spatial image intensity information to intensity-only considerations, specifically those associated with the intensity histogram. Although significantly simplifying computational processing, this transformation results in the loss of important spatial cues for identifying salient imaging features, such as ventilation defects, which have been identified as correlating with lung pathophysiology. In this work, we demonstrate the interrelatedness of the most common approaches for histogram-based, ventilation segmentation of hyperpolarized gas lung imaging for driving voxelwise classification. We evaluate the underlying assumptions associated with each approach and show how these assumptions lead to suboptimal performance. We then illustrate how a convolutional neural network can be constructed in a multi-scale, hierarchically feature-based (i.e., spatial) manner which circumvents the problematic issues associated with existing intensity-only approaches. Importantly, we provide the entire evaluation framework, including this newly reported deep learning functionality, as open-source through the well-known Advanced Normalization Tools (ANTs) library.

Introduction

Early acquisition and development

Early hyperpolarized gas pulmonary imaging research reported findings in qualitative terms.

Descriptions:

- "^{3}He MRI depicts anatomical structures reliably" (1)
- "hypointense areas" (2)
- "signal intensity inhomogeneities" (2)
- "wedge-shaped areas with less signal intensity" (2)
- "patchy or wedge-shaped defects" (3)
- "ventilation defects" (4)
- "defects were pleural-based, frequently wedge-shaped, and varied in size from tiny to segmental" (4)

Historical overview of quantification

Initial attempts at quantification of ventilation images were limited to ennumerating the number of "ventilation defects" or estimating ventilation defect percentage (as a percentage of total lung volume). Often these measurements were acquired on a slice-by-slice basis.

Prior to the popularization of deep learning in medical image analysis, including in the field of hyperpolarized gas imaging (5), widely used semi-automated or automated segmentation techniques were primarily based on intensity-only considerations. In order of increasing sophistication, these techniques can be categorized as follows:

- binary thresholding based on relative intensities (6),

- linear intensity standardization based on global rescaling of the intensity histogram to a reference distribution based on healthy controls, i.e., "linear binning" (7, 8),
- non-linear intensity standardization based on piecewise affine transformation of the intensity histogram using the K-means algorithm (9), and
- Gaussian mixture modeling (GMM) with spatial constraints using Markov random field (MRF) modeling (10).

The early semi-automated technique used to compare smokers and never-smokers in (6) uses manually drawn regions to determine the mean signal intensity as well as the standard deviation of the noise to derive a threshold value of three noise standard deviations below the mean intensity. All voxels above that threshold value were considered "ventilated" for the purposes of the study. Similar to the histogram-only algorithms (i.e., linear binning and k-means), this approach does not take into account the various artefacts associated with MRI such as the non-Gaussianity of the MR imaging noise (11, 12) and the intensity inhomogeneity field (13).

To provide a more granular categorization of ventilation that tracks with clinical qualitative assessment, an increase in the number of voxel classes have been added to the various lung parcellation protocols beyond the binary categories of ventilated and non-ventilated. Linear binning is a simplified intensity standardization approach with six discrete intensity levels (or clusters). The six clusters are evenly spaced throughout the intensity range based on the mean and standard deviation values determined from a cohort of healthy controls all rescaled to $[0, 1]$. Such rescaling for determination of segmentation clusters of lung images in a particular study can be thought of as a global affine 1-D transform of the intensity histogram. Note that such a global transform does not account for MR intensity non-linearities that have been well-studied (14–17) and can cause significant intensity variation even in the same subject due to a variety of conditions. As stated in (18):

> Intensities of MR images can vary, even in the same protocol and the same sample and using the same scanner. Indeed, they may depend on the acquisition conditions such as room temperature and hygrometry, calibration adjustment,

slice location, B0 intensity, and the receiver gain value. The consequences of intensity variation are greater when different scanners are used.

As we demonstrate in subsequent sections, ignoring these non-linearities can have significant consequences in the well-studied (and somewhat analogous) area of brain tissue segmentation in T1-weighted MRI (e.g., (19–21)).

Finally, we point out that N4 bias correction is used in many of these algorithms which is also histogram-based.

It should be noted that we are not claiming that these algorithms are erroneous. Much of the relevant research has been limited to quantifying differences with respect to ventilation vs. non-ventilation in various clinical categories and these algorithms have certainly demonstrated the capacity for advancing such research. However, as acquistion and analyses methodologies improve, so should the level of sophistication and performance of the measurement tools.

Methods

Results

Discussion

References

1. Bachert P, Schad LR, Bock M, et al.: Nuclear magnetic resonance imaging of airways in humans with use of hyperpolarized 3He. *Magn Reson Med* 1996; 36:192–6.

2. Kauczor HU, Hofmann D, Kreitner KF, et al.: Normal and abnormal pulmonary ventilation: Visualization at hyperpolarized he-3 mr imaging. *Radiology* 1996; 201:564–8.

3. Kauczor HU, Ebert M, Kreitner KF, et al.: Imaging of the lungs using 3He mri: Preliminary clinical experience in 18 patients with and without lung disease. *J Magn Reson Imaging*; 7:538–43.

4. Altes TA, Powers PL, Knight-Scott J, et al.: Hyperpolarized 3He MR lung ventilation imaging in asthmatics: Preliminary findings. *J Magn Reson Imaging* 2001; 13:378–84.

5. Tustison NJ, Avants BB, Lin Z, et al.: Convolutional neural networks with template-based data augmentation for functional lung image quantification. *Acad Radiol* 2019; 26:412–423.

6. Woodhouse N, Wild JM, Paley MNJ, et al.: Combined helium-3/proton magnetic resonance imaging measurement of ventilated lung volumes in smokers compared to never-smokers. *J Magn Reson Imaging* 2005; 21:365–9.

7. He M, Driehuys B, Que LG, Huang Y-CT: Using hyperpolarized 129Xe mri to quantify the pulmonary ventilation distribution. *Acad Radiol* 2016; 23:1521–1531.

8. He M, Wang Z, Rankine L, et al.: Generalized linear binning to compare hyperpolarized 129Xe ventilation maps derived from 3D radial gas exchange versus dedicated multislice gradient echo mri. *Acad Radiol* 2020; 27:e193–e203.

9. Kirby M, Heydarian M, Svenningsen S, et al.: Hyperpolarized 3He magnetic resonance functional imaging semiautomated segmentation. *Acad Radiol* 2012; 19:141–52.

10. Tustison NJ, Avants BB, Flors L, et al.: Ventilation-based segmentation of the lungs using hyperpolarized (3)He MRI. *J Magn Reson Imaging* 2011; 34:831–41.

11. Gudbjartsson H, Patz S: The rician distribution of noisy mri data. *Magn Reson Med* 1995; 34:910–4.

12. Andersen AH: On the rician distribution of noisy mri data. *Magn Reson Med* 1996; 36:331–3.

13. Sled JG, Zijdenbos AP, Evans AC: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging* 1998; 17:87–97.

14. Wendt RE 3rd: Automatic adjustment of contrast and brightness of magnetic resonance images. *J Digit Imaging* 1994; 7:95–7.

15. Nyúl LG, Udupa JK: On standardizing the mr image intensity scale. *Magn Reson Med* 1999; 42:1072–81.

16. Nyúl LG, Udupa JK, Zhang X: New variants of a method of mri scale standardization. *IEEE Trans Med Imaging* 2000; 19:143–50.

17. De Nunzio G, Cataldo R, Carlà A: Robust intensity standardization in brain magnetic resonance images. *J Digit Imaging* 2015; 28:727–37.

18. Collewet G, Strzelecki M, Mariette F: Influence of mri acquisition protocols and image intensity normalization methods on texture classification. *Magn Reson Imaging* 2004; 22:81–91.

19. Zhang Y, Brady M, Smith S: Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 2001; 20:45–57.

20. Ashburner J, Friston KJ: Unified segmentation. *Neuroimage* 2005; 26:839–51.

21. Avants BB, Tustison NJ, Wu J, Cook PA, Gee JC: An open source multivariate framework for n-tissue segmentation with evaluation on public data. *Neuroinformatics* 2011; 9:381–400.