The Journal of Open Source Software

mednoise: a python package for medical image processing

Ravi Bandaru[1]

1 Research Assistant, Cincinnati Children's Hospital - Division of Human Genetics

DOI: 10.21105/joss.03597

Software

- Review
- Repository
- Archive

Editor: Øystein Sørensen

Reviewers:

- @kasbohm
- @ethanknights

Submitted: 12 July 2021
Published: 10 August 2021

Summary

mednoise is a library for processing medical images in Python. It contains various algorithms which are highly tuneable, enabling it to encompass a wide variety of scientific use-cases. The algorithms contained in mednoise are grounded in NumPy arrays (Harris et al., 2020). mednoise treats each image as a two-dimensional array of pixels which is processed with a variety of manipulations and substitutions.

When given a series of medical images, mednoise runs user-requested computations, eventually yielding a final medical image, without noise, that is primed to be analyzed by machine learning and deep learning models.

These noiseless images improve the accuracy of AI imaging models, both supervised and unsupervised, significantly (Karimi et al., 2020). mednoise offers an easy-to-use terminal interface while processing medical images. Additionally, its inference-compatible nature allows for easy integration into existing scientific workflows, like Snakemake (Mölder et al., 2021). By processing and outputting medical images that heavily improve image-specific AI models, mednoise facilitates high-speed scientific model creation and image analysis.

Statement of need

With the field of artificial intelligence and bioinformatics growing almost exponentially, programming languages have adapted to include more machine learning and deep learning libraries. Within Python, there is TensorFlow, a module that offers the ability for users to develop AI models with minimal keystrokes (Developers, 2021).

The field of medical imaging, however, has not adapted with similar speed. Medical imaging tools used in medical offices across the world are rarely new and are not optimized for modern AI models (Lundervold & Lundervold, 2019).

Enter mednoise, an easily integrable Python tool designed to process medical image metadata into usable formats for machine learning models.

More specifically, mednoise removes unnecessary noise that has unremarkable clinical significance, yielding medical images that have the highest impact possible in bioinformatics research.

Many machine learning and deep learning models that analyze images do so by undergoing a simple computational process – converting an image into a series of pixels, and analyzing those pixels in relation to some structure (ex. a specified diagnosis) to find an inferred structure through mathematical formulas.

Inevitably, images have some *noise*, areas of pixels that have little to no clinical significance. This could be letters, numbers, and identifying shapes that help align the image and identify it to the patient. To researchers, these pixels hinder the accuracy of their machine learning models as they contain color-specific pixel information that detracts from the areas of significance, either severely inhibiting the accuracy of models or artificially inflating their accuracy (if they contain diagnosis-related structure within them)(Karimi et al., 2020).

mednoise uses a variety of algorithms to reduce and often eliminate noise from images. Naturally, medical images are often varied and have little to no order to them. mednoise is built under that assumption, placing the power to control noise reduction directly in the hands of the researcher.

The design of mednoise lends the package immense potential. mednoise can be easily integrated into any workflow, high-performance computing network, and even medical imaging tools themselves. With its open and highly tuneable structure, this package is useful to almost any usecase related to medical image research.

mednoise is designed as a tool for pipelines and workflow management systems that rely on medical image metadata (or at least the analysis of that metadata). It aims to create a future of medical imaging that is more accessible and utilizable in the constantly evolving field of bioinformatics.

Example

The following example demonstrates the processing of an ultrasound image with mednoise, specifically its branch algorithm. The example makes use of the example image, hosted on the documentation website, that is provided for generation of reproducible bug reports.

```
>>> from mednoise import branch as md
>>> md.branch_complete("/example/directory/file.PNG", 415, 400,
    iterations = 350)
#file.PNG contains example image from mednoise.github.io/exampleimage.html
md.branch_complete - Image 1 Importing:0:00:01
md.branch_complete - Image 1 Converting:0:00:00
md.branch_complete - Image 1 Translating:0:00:00
md.branch_complete - Image 1 Branching:0:04:04
md.branch_complete - Image 1 Branch Analyzing:0:01:16
md.branch_complete - Image 1 Branch Isolating:0:00:02
md.branch_complete - Image 1 Array Priming:0:00:00
md.branch_complete - Image 1 Translating:0:00:00
md.branch_complete - Image 1 Saving:0:00:00
```

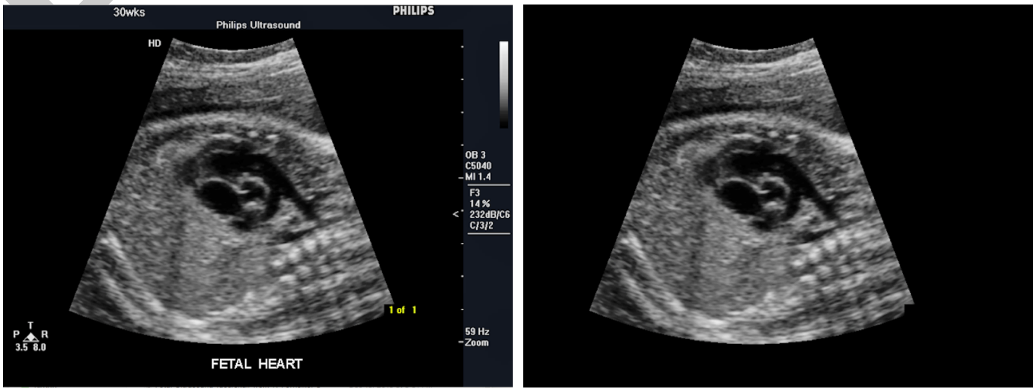

Figure 1: An example usage of mednoise with the input file (left) being silenced by the branch_complete algorithm yielding the final, primed image (right)

References

Developers, T. (2021). *TensorFlow* (Version v2.6.0-rc0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.5043456

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Karimi, D., Dou, H., Warfield, S. K., & Gholipour, A. (2020). Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Medical Image Analysis*, *65*, 101759. https://doi.org/10.1016/j.media.2020.101759

Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik*, *29*(2), 102–127. https://doi.org/10.1016/j.zemedi.2018.11.002

Mölder, F., Jablonski, K., Letcher, B., Hall, M., Tomkins-Tinch, C., Sochat, V., Forster, J., Lee, S., Twardziok, S., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). Sustainable data analysis with snakemake [version 1; peer review: 1 approved, 1 approved with reservations]. *F1000Research*, *10*(33). https://doi.org/10.12688/f1000research.29032.1