Analyzing Figures of Brain Images from Alzheimer's Disease Papers

Satoshi Tsutsui[1], Guilin Meng[2], Xiaohui Yao[3], David Crandall[1], Ying Ding[1,4,5]
[1]School of Informatics and Computing, Indiana University, Indiana, USA
[2]Department of Neurology, Tenth People's Hospital, Tongji University, Shanghai, China
[3]Department of Neurology, Indiana University Center for Neuroimaging, Indiana, USA
[4]Library, Tongji University, Shanghai, China
[5]School of Information Management, Wuhan University, Wuhan, China

Abstract
Which papers focusing on Alzheimer's disease (AD) include MRI scans of human brains? These images play an important role in clinical detection of AD, but finding them currently requires manual inspection of papers after a keyword search. In order to provide AD researchers with a more efficient way of finding relevant papers, here we focus on three preliminary problems involving automatically identifying figures containing brain images, and solve them as automatic image classification tasks. This is a first step towards efficiently allowing AD researchers to retrieve papers containing a particular type of brain image (e.g. of a patient). We report preliminary results from a larger project, in collaboration with AD researchers.

Keywords: Alzheimer's Disease; Figure Mining; Viziometrics

Contact: stsutsui@indiana.edu.

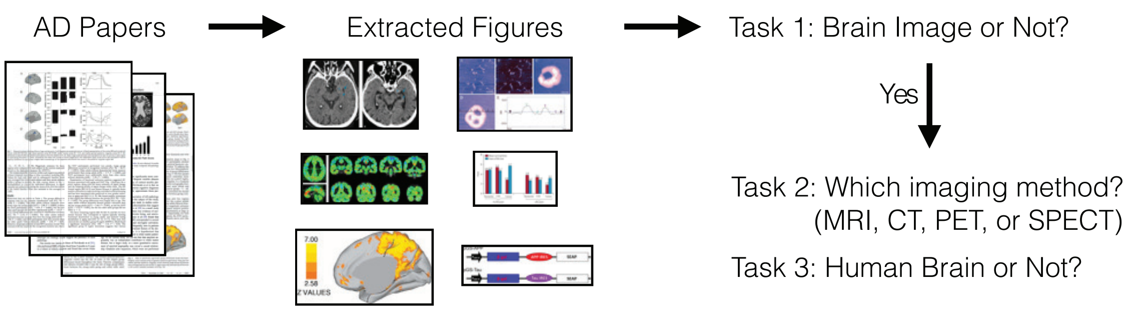

Figure 1: In this paper, we show how to extract figures from PDFs of AD papers, and answer three questions: Is it a brain image or not? What kind of imaging technology is used? Is it a human brain or not?

1 Introduction

The exponential rate of growth of biomedical literature (Larsen & Von Ins, 2010) makes it hard for researchers to follow the latest progress even in their specialized domain. Brain imaging researchers, for example, must manually inspect large numbers of papers, even after performing a text-based search, to find those containing brain images. With that issue in mind, we currently collaborate with researchers on Alzheimer's disease (AD), aiming to contribute to AD research from computer and information science perspective.

Brain imaging plays a key role for clinical detection of AD. The most widely used biomarkers of AD include focal atrophy and metabolic dysfunction, which can be detected by brain imaging. In fact, a definite diagnosis of AD requires the examination of brain tissue, which is usually performed after the death of a patient (Ballard et al., 2011). This fact shows that brain imaging, imaging, a way of inspecting a brain without surgery, plays a key role for clinical detection of AD.

In this paper, we focus on analyzing brain images in AD papers. Currently, the purpose is to assist AD researchers to find papers containing a brain image of their interest, but it also has a potential for assisting clinical diagnosis because we could retrieve papers containing a brain image that is similar to that of a patient of interest.

This paper reports preliminary results from a larger project collaborating with AD researchers. We investigate figures of brain images, and identify three specific problems to solve (Section 3). Then we regard these as image classification problems, and approach them using computer vision techniques (Section 4). We also discuss several future challenges (Section 5).

2 Related Work

Analyzing figures in scientific papers has recently become an interesting area of research. Several works (Choudhury, Mitra, & Giles, 2015; Clark & Divvala, 2015; Kuhn, Luong, & Krauthammer, 2012; Clark & Divvala, 2016) extract figures directly from PDFs. In addition to extraction, there are lines of works (Savva et al., 2011; Choudhury & Giles, 2015; Siegel, Horvitz, Levin, Divvala, & Farhadi, 2016) that classify figures. However, they focus on classification into general categories such as diagrams, photos, tables or data plots. In contrast, we are interested particularly in brain imaging in the AD domain. In addition to focusing on individual figures, Viziometrics (Lee, West, & Howe, 2016a, 2016b) focus on analysis of figures from collections of papers. It has recently been proposed emphasizing the visual content analysis of papers in contrast to bibliometrics, which has been focused only on textual content of papers.

3 Problem Specifications

We consulted with an AD expert to investigate what kinds of information are important to them when looking at figures of brain images. We manually checked 30 random AD papers, examined the figures in them, and defined the following problems as preliminary steps:

1. Does the figure contain an image of a brain?
2. If so, which imaging methods was used to capture it? (e.g. Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT))
3. Is it a human brain or not (e.g. a rat)?

4 Computer Vision Approach

The three questions above can be posed in terms of image classification, which is well-studied in computer vision. The corresponding classification problems are:

Task 1 Given a figure, classify whether it contains brain images or not.
Task 2 Given a figure containing brain images, classify into four imaging types: MRI, CT, PET, SPECT
Task 3 Given a figure containing brain images, classify whether it is a human brain or not.

It is possible to encounter figures that have multi-labels, but in this preliminary study, we excluded such cases.

4.1 Dataset Preparation

The state-of-the-art image classification approach is to use deep neural networks (LeCun, Bengio, & Hinton, 2015; Razavian, Azizpour, Sullivan, & Carlsson, 2014; He, Zhang, Ren, & Sun, 2015). Hence we need to prepare training data. We manually downloaded PDF files of 100 articles via PubMed search. We first searched with the keywords *Alzheimer* and *Brain Imaging*, and randomly downloaded 50 articles which we expected to contain a variety of brain images. Then, in order to get negative examples (i.e. figures not containing brain images), we also downloaded another 50 articles by just searching *Alzheimer*.

After obtaining 100 PDFs, we extracted 302 figures using PDFFigures 2.0 (Clark & Divvala, 2016). We manually selected 66 figures containing brain images, for use as positive examples for task 1. For negative examples, we randomly extracted 66 figures from the remaining ones that do not contain brain images. This yielded 132 figures in total. Some examples are shown in Figure 2 and 3.

For the 66 figures containing brain images, we investigated their types of brain images. In fact, the figures are often composed of multiple brain images of different types, and even non-brain imaging figures such as plots. We first manually (multi-)labeled each figure into MRI, PET, CT, and SPECT. This labeling requires some expertise, and took a few hours for an AD expert because sometimes the images are post-processed (e.g. coloring or annotation), which makes it hard to confidently distinguish the imaging method and requires the expert to read the contents of the paper. This labeling resulted in the data for task 2 (38 MRI images, 27 PET images, 4 CT images, and 4 SPECT images). When training a classifier, we only used images having a single label. Also, we annotated whether the brain is human or not, forming the dataset for task 3. Of the 66 brain images, 12 figures are non-human brains. We did not find figures containing brain images of both humans and animals.

When training classifiers, we use equal numbers of data for each label to avoid biasing the classifiers. In other words, we limited the training data by the least frequent label. As a result, we only have 132 images for task 1, 16 for task 2, and 24 for task 3. The dataset is available on the web [1].

4.2 Figure Classification Approach

The state-of-the-art approach in image classification uses deep learning (LeCun et al., 2015), specifically convolutional neural networks. However, it requires much more training instances than we collected if we train from scratch, so we used a transfer learning approach (Razavian et al., 2014). We used a pre-trained 50 layer deep residual network (ResNet50) (He et al., 2015) on 1.2 million images from ImageNet (Russakovsky et al., 2015), and re-trained only the last classification layer. We note that it might be possible to use more complicated approach to obtain a better performance, but the focus on this preliminary paper is not optimizing the performance.

4.3 Classification Results

Since we only have an limited amount of data (especially for task 2), we evaluated the accuracy using 4-fold cross validation. Recognizing brain images works relatively well (93 % and 83 % accuracy on task 1 and 3, respectively, compared to a 50% baseline of random guessing). However, recognizing imaging type does not work as well (56 % accuracy on task 2), although this is a more difficult task (25% random baseline). In order to understand the difficulties, we compared an aggregated confusion matrix shown in Table 1. It indicates that CT is always classified correctly, however, PET and SPECT are difficult to distinguish. We note that the size of data is a limitation of our study, especially for task 2 with only 16 examples. Hence it might not be valid to generalize our results. We discuss this limitation in the next section.

		Predicted			
		MRI	PET	CT	SPECT
Truth	MRI	1	1	2	0
	PET	0	2	0	2
	CT	0	0	4	0
	SPECT	0	2	0	2

Table 1: Confusion matrix for brain image type recognition (task 2)

[1] http://homes.soic.indiana.edu/stsutsui/ad_brain_images/

5 Challenges and Future Work

We noticed two challenges. First, the cost of collecting training data is quite high. State-of-the-art computer vision techniques require large amount of data. Crowdsourcing (e.g. Amazon Mechanical Turk) is often used in computer vision research to annotate images with ground truth labels, but annotating medical figures requires domain expertise, so crowdsourcing is difficult to use. It is not practical to ask many AD researchers to annotate. A possible solution is to annotate examples in a semi-automatic manner using some heuristics (Mintz, Bills, Snow, & Jurafsky, 2009). It might be possible to use figure captions to automatically; however, these annotations always have noise, so we need techniques with noise robustness.

Another challenge is to localize brain images within a figure because many figures have multiple types of brain images. In computer vision, techniques based on Region-based Convolutional Neural Networks (R-CNN) (Girshick, Donahue, Darrell, & Malik, 2014) are the state-of-the-art. However, they again require large amounts of large training data. Some previous work in figure mining (Siegel et al., 2016; Lee & Howe, 2015) parse figures into sub-figures, but they assume plots or charts with a white background, so these techniques cannot be directly applied to brain images whose background is sometimes black (e.g. right bottom of Figure 2).

6 Conclusion

This study presents preliminary results from a project collaborating with AD researchers. In this paper, we focused on brain image analysis from AD papers. As an initial step, we identified three specific problems, and discussed future challenges. However, of course, many more problems remain to be solved other than these three, such as recognizing which region of a brain is highlighted in a brain image, or segmenting brain images out of complex multi-part figures

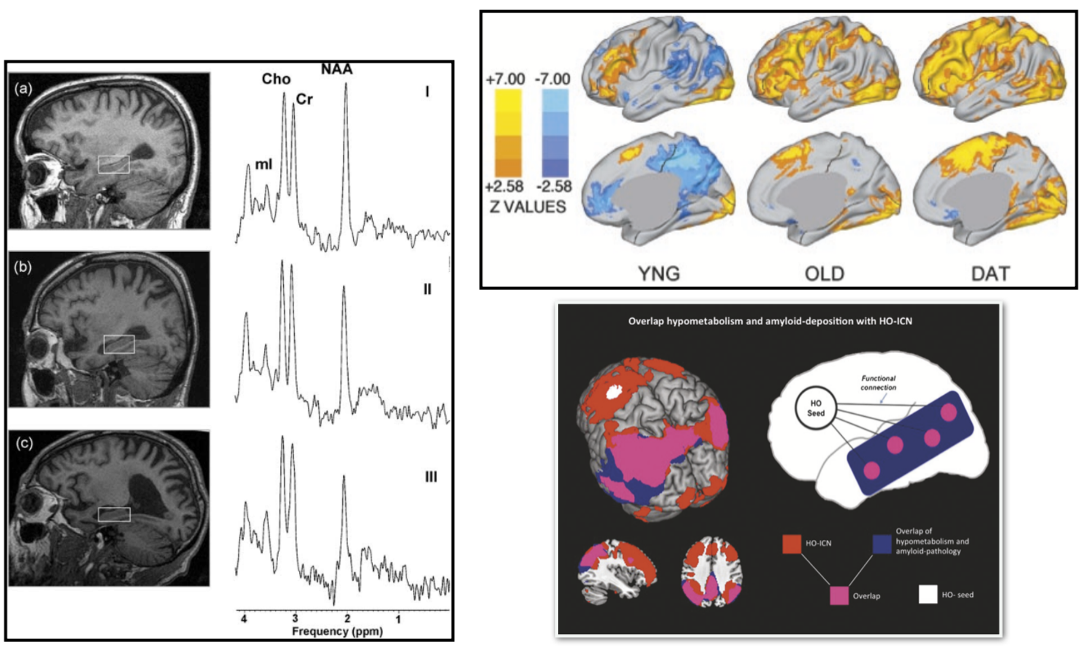

Figure 2: Example figures containing brain images. (Figure sources PMID: 15905028, 14608034, 24870443)

References

Ballard, C., Gauthier, S., Corbett, A., Brayne, C., Aarsland, D., & Jones, E. (2011). Alzheimer's disease. *The Lancet*, *377*(9770), 1019 - 1031. Retrieved from http://www.sciencedirect.com/science/article/pii/S0140673610613499 doi: http://dx.doi.org/10.1016/S0140-6736(10)61349-9

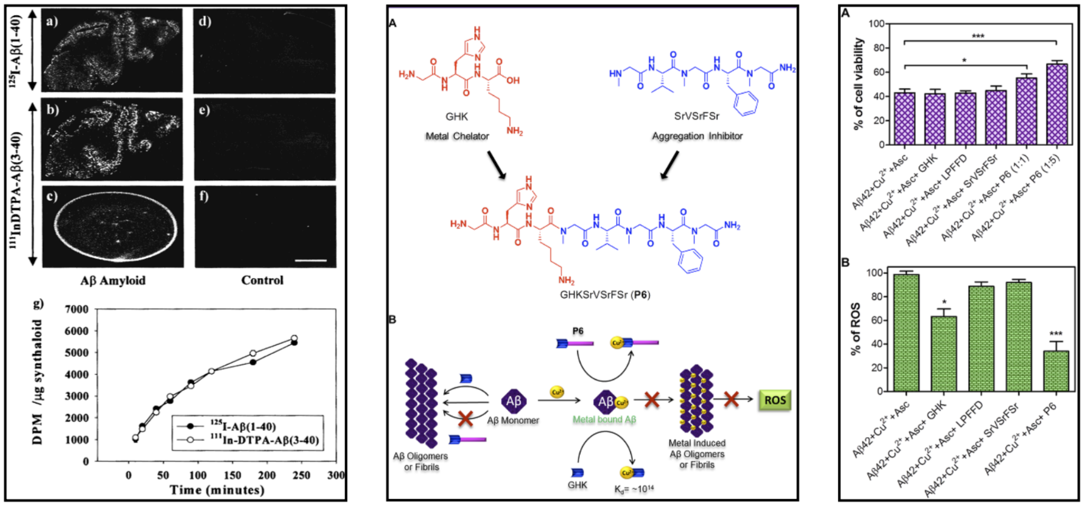

Figure 3: Example figures not containing brain images. (Figure sources PMID: 11906265, 27355515)

Choudhury, S. R., & Giles, C. L. (2015). An Architecture for Information Extraction from Figures in Digital Libraries. In *International World Wide Web Conference (WWW).*

Choudhury, S. R., Mitra, P., & Giles, C. L. (2015). Automatic extraction of figures from scholarly documents. In *Proceedings of the 2015 ACM Symposium on Document Engineering* (pp. 47–50). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2682571.2797085 doi: 10.1145/2682571.2797085

Clark, C., & Divvala, S. (2015). Looking Beyond Text: Extracting Figures, Tables, and Captions from Computer Science Paper. In *Association for the Advancement of Artificial Intelligence (AAAI) Workshop on Scholarly Big Data.*

Clark, C., & Divvala, S. (2016). PDFFigures 2.0: Mining Figures from Research Papers. In *Joint Conference on Digital Library (JCDL).*

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Computer Vision and Pattern Recognition (CVPR).*

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *arXiv preprint arXiv:1512.03385*.

Kuhn, T., Luong, T., & Krauthammer, M. (2012). Finding and accessing diagrams in biomedical publications. In *Association for the Advancement of Artificial Intelligence (AAAI).*

Larsen, P. O., & Von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, *84*(3), 575–603.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Lee, P., & Howe, B. (2015). Dismantling Composite Visualizations in the Scientific Literature. In *International Conference on Pattern Recognition Applications and Methods (ICPRAM).*

Lee, P., West, J., & Howe, B. (2016a). Viziometrics: Analyzing Visual Patterns in the Scientific Literature. *Journal of the Association for Information Science and Technology (JASIST) (in prep)*.

Lee, P., West, J. D., & Howe, B. (2016b). VizioMetrix: A Platform for Analyzing the Visual Information in Big Scholarly Data. In *International World Wide Web Conference (WWW).* doi: 10.1145/2872518.2890523

Mintz, M., Bills, S., Snow, R., & Jurafsky, D. (2009). Distant Supervision for Relation Extraction Without Labeled Data. In *Association for Computational Linguistics (ACL).*

Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. In *Computer Vision and Pattern Recognition (CVPR) Workshops.*

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, *115*(3), 211-252. doi: 10.1007/s11263-015-0816-y

Savva, M., Kong, N., Chhajta, A., Fei-Fei, L., Agrawala, M., & Heer, J. (2011). Revision: Automated classification, analysis and redesign of chart images. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 393–402).

Siegel, N., Horvitz, Z., Levin, R., Divvala, S., & Farhadi, A. (2016). FigureSeer: Parsing Result-Figures in Research Papers Computer Vision for Scholarly Big Data. In *European Conference on Computer Vision (ECCV).*