Example 1:

Paper Review: Is a PET All You Need? A Multi-modal Study for Alzheimer's Disease Using 3D CNNs

1. SUMMARY

This paper performs an extensive study on two different diagnosis techniques for classifying Alzheimer's Disease (AD): structural MRI (sMRI) and fluorodeoxyglucose positron emission tomography (FDG-PET). Specifically, it looks into the merits of a single vs multi-modal DNN approach using both sMRI and FDG-PET for classification. Both modalities are tested first individually and then in tandem using three multi-modal fusion techniques: early, middle, and late. The results are evaluated with a comparison of the mean balanced accuracy (BACC) and standard deviation across the 3 fusion models and on 3 evaluation schemes of the input data: correct, random MRI, and random PET to test the relevance of MRI in a multi-modal approach. The results consistently showed the multi-modal approach under-performed both the single-modal FDG-PET and MRI, with single FDG-PET outperforming MRI. A further study was also conducted to show the extent to which MRI and FDG-PET contributed to overall predictions and showed that PET contributed 1.77 times more than MRI. Narazani et al., contributions lie in their extensive study of multi-modal DNN architecture vs single-modal approaches for predictions and stress future research look into opportunities for conformity of multi-modal techniques in clinical practice, and better assessment of single-modality in models.

2. STRENGTHS

- The author's post-hoc explanation of relevance maps is helpful to understand not only which modality contributed more to a given prediction, but successfully quantifies by *how much*. It allows the reader to have a better understanding of the extent to which PET plays a role instead of using the BACC metric for all justifications.

- The three fusion strategies used in the study do a good job to account for a multi-modal architecture and are more holistic for justifying the results of this study. Because there are different ways to incorporate multi-modal information in a CNN, trying out different fusion strategies accounts for how different ways of sharing data between the 2 modals can impact performance. Having 3 separate sharing mechanisms and testing them is a way to strongly justify the ineffectiveness of the multi-modal architecture for the classification task.

- Figure 1 is very effective at communicating the 3 fusion strategies used in the experiment. It enhances our understanding of how the 2 threads share information over the period of the network and how they compare.

3. WEAKNESSES

- In the section explaining the randomly exchanged image pairs for training(page 73), there is limited explanation on the need for PET pre-processing. While the author does cite a study showing decreased performance for un-processed PET data by 7% on performance, some insight into PET training strategies would be useful.

- The dataset used for training of the images for AD, CN, and MCI vary quite a bit by size. The AD dataset is only for 257 patients, while the MCI dataset has almost 3 times as much data with 611 patients. A bigger dataset for example can be prone to more over-fitting and decreased variance. The datasets should try to minimize model parameter bias by keeping datasets as consistent in size and quality as possible.

- When studying the results of Table 1 and comparing the performance of middle and late fusion models for random data, the author makes mention of a drop of performance in FDG-PET but there is missing substantive reasoning for this causality. A loss of information is inferred as the reason for the decreased performance but it is ambiguous.

4. TECHNICAL EXTENSIONS

- It is worth exploring different multimodal procedures referenced in this paper to see if results are similar to those of a combined MRI and FDG-PET model. Like CSF, metadata on demographics, sex and examination results on cognitive abilities. There is an emphasis on using clinician practical modalities and understanding each single modality model in more depth. Having more of these multi-modals will add more data points to AD prediction abilities.

- Re-doing the study of single vs multi-modal models in FDG-PET and MRI for solving a different classification problem other than AD would be interesting to compare results. The current understanding of results is in the context of prediction of AD, however testing these results on different prediction problems would give a better idea of how context driven this approach is.

5. OVERALL REVIEW

This paper performs a thorough investigation of multi-modality models vs single modal models for prediction of AD. While some substantive reasoning may be lacking in explanation for BACC performance drop off in randomization of PET and MRI images, the paper makes a strong case for better understanding single-modalities in future research. On the NIPS review scale, I'd rate this paper 1. The experimental design is very systematic and baselines are well established. Explanations for the 3 fusion strategies are all done and do a good job at holistically testing the multi-modality approach. Future directions are well understood and proper acknowledgements were made to supporting institutions.

Example 2:

Paper Review: - Is a PET all you need? A multi-modal study for Alzheimer's disease using 3D CNNs

1. Summary

This paper investigates the results of recent works in neuroimaging-based computer-aided Alzheimer's disease (AD) diagnosis with DNNs, which claimed that fusing sMRI and FDG-PET leads to improved accuracy in AD classification. The authors' stated main objective is to "rigorously evaluate whether MRI is truly relevant for diagnosing AD when FDG-PET is available too".

The authors propose an evaluation framework for deep learning for modality fusion, and use this as the baseline for their experiments. They perform an ablation study that evaluates the performance of different fusion techniques; they examine a binary classification (AD vs. control) and three-class classification (AD vs. mild cognitive impairment vs. control).

The authors suggest that their results demonstrate that a single-modality network using FDG-PET performs best for healthy/AD classification, and argue for 1) the use of their framework and 2) taking existing clinical knowledge (e.g. for AD biomarkers) into account for future multi-modal studies.

2. Strengths

- Uses established clinical knowledge as a baseline to question recent experimental results in deep learning.
- Strong evidence for their claim that a PET is all you need.
- Good use of figures: 1) clear model architecture diagram, 2) visual explanation of different fusion strategies, 3) relevance maps to show what parts of image are contributing to predictions (similar to class activation maps).
- Clear writing. The main idea, their contribution, and their experimental setup are all very clear.
- They use a single CNN architecture for all of their experiments, which makes it clear that the only difference in performance is due to the choice of inputs to the multi-modal fusion approach (early vs. middle vs. late fusion and correct vs. random MRI vs random PET).
- The authors claim to be the first to use channel exchange for middle fusion on AD images.

3. Weaknesses

- Repetitive writing. They repeat the idea that FDG-PET better captures AD-specific patterns of neurodegeneration than MRI many separate times (though one of the occurences is in the abstract). Though repetitive, it makes the idea very clear.
- The paper is primarily experimental, and applies existing techniques/methods. There is not much theoretical novelty.
- The authors do not open-source their code or data, which hinders reproducibility. Their method is clear and could be reimplemented, but it puts the burden on the reader.
- There are some typos throughout the paper. Example (last sentence of Methods intro): "This allows us to quantify **to** importance of each modality".
- The authors only evaluate their proposed framework for evaluating the contribution of multimodal data fusion on a single dataset. Their proposal for how to adapt their framework to other datasets is clear, but they do not demonstrate this on more than one dataset.
- The authors use a CNN framework from 2015; SOA methods may produce better results on MRI data.

4. Concluding Remarks

Overall, I would give this paper a rating of 6 (NIPS), borderline accept. They have a strong conclusion and propose a useful framework, but do not offer new methods/models/theory and only evaluate on a single dataset.

Example 3:

Paper#1 Review: Is a PET All You Need? A Multi-modal Study for Alzheimer's Disease Using 3D CNNs

1. Summary

- The paper explores the use of 3D convolutional neural networks (CNNs) for diagnosing Alzheimer's disease using positron emission tomography (PET) scans.
- The 3D CNNs were trained on a dataset that included both PET and structural magnetic resonance imaging (MRI) scans.
- The results showed that the 3D CNNs could accurately diagnose Alzheimer's disease based on PET scans alone, with an accuracy of 89.1
- The study provides a new approach to Alzheimer's disease diagnosis and highlights the potential of using AI and deep learning techniques in medical imaging.
- The authors have made the code used in the study available to the public, enabling further research and validation of the results.

2. Strengths

- Innovative approach: The study provides a novel approach to the diagnosis of Alzheimer's disease, using a combination of positron emission tomography (PET) scans and deep learning techniques.
- High accuracy: The results of the study showed that the 3D convolutional neural networks (CNNs) used in the study were able to accurately diagnose Alzheimer's disease based on PET scans alone, with a high accuracy of 89.1%.
- Relevance to clinical practice: The results of the study have the potential to be applied in clinical practice and improve the accuracy of Alzheimer's disease diagnosis.
- Open-source code: The authors have made the code used in the study available to the public, which allows other researchers to build upon the work and validate the results.

3. Weaknesses

- Limited use of other imaging modalities: Although the study used both PET and MRI scans, it did not explore the use of other imaging modalities that are commonly used in the diagnosis of Alzheimer's disease, such as computed tomography (CT) scans.
- Sample size: The study was conducted on a relatively small sample of patients, and the results may not be generalizable to larger populations.
- Limited diversity of the sample: The study was conducted on a homogeneous sample, and it is not clear how well the results would generalize to more diverse populations.
- Validation: The results of the study were not independently validated. Further studies with larger sample sizes and independent validation are needed to confirm the results.
- Transferability: The 3D CNNs developed in this study were trained and tested on a specific dataset, and it is not clear how well they would perform on other datasets or in different populations.

4. Ideas for future research

- The proposed work used 3D-CNN to understand the PET data, however, it is interesting to utilize a full self-attention or LongFormer [1] kind of model to exploit the non-local affinities of the data.
- Instead 3 types of fusion strategies are mentioned in Fig. 1 of the paper, it is interesting to explore some kind of cross-modal contrastive regularization objective or using cross-attention across random PET and MRI images.

References

[1] Iz Beltagy, Matthew E Peters, and Arman Cohan. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*, 2020. 1