Response to the Reviews and Decision

Title: Interactive Deep Learning for Investigative Sorting of Plant Images by Visual Phenotypes

DOI:
10.1002/essoar.10508768.1

Authors:
Huimin Han
Ritvik Prabhu
Timothy Smith
Kshitiz Dhakal
Xing Wei
Song Li
Chris North

Date: Janunary 26, 2022

Message from the Authors

Dear Editors and Reviewers,

We thank you for your constructive comments, which are valuable and very helpful for revising and improving our paper. We have addressed the comments and incorporated your valuable suggestions in the revised version. The updated contents are colored in red in the revised paper to differentiate with contents in the original one.

We address each comment separately in the following detailed response. The comments we received are boxed, and our responses are written following each comment. We've added more clarification about the relationship between our approach and transfer learning, the validation for our approach, and modelling choices. We hope that you find the revised version satisfactory.

Sincerely,
Huimin Han, Ritvik Prabhu, Timothy Smith, Kshitiz Dhakal, Xing Wei, Song Li, Chris North.

Response To Reviewer #1

Overall Comments

> "The writing is clear although some statements are too general and/or no justification is provided. For example, the authors claim that they can change the classification task of a DNN and reuse a previous DNN as the 'majority of parameters in the lower levels of the neural networks do not need to be retrained' but provide no evidence or citation to support this. They appear to be suggesting an approach to transfer learning but never mention this term or related work in this area. They provide no classification results so it is unclear whether the extracted features are useful. Yes, input from the user can suggest new feature weights, but how this impacts classification accuracy is not discussed."

Response

We appreciate your careful review and detailed feedback. Our focus in the revised version was to clearly state the relationship between our approach and transfer learning and the clarification for the usefulness of the extracted features. We hope that you find the following response satisfactory.

Reviewer Comment

> "They appear to be suggesting an approach to transfer learning but never mention this term or related work in this area."

Response

Thanks for your review and yes we used a pre-trained model as a feature extractor, which is one of the three major Transfer Learning scenarios[1], we've added more reference papers about using pre-trained model as a feature extractor in the revised manuscript.

Reviewer Comment

> "They provide no classification results so it is unclear whether the extracted features are useful. Yes, input from the user can suggest new feature weights, but how this impacts classification accuracy is not discussed."

Response

Thank you for the comment. Talking about those extracted features, the main process we did is to take a CNN pretrained on ImageNet, remove the last fully-connected layer (this layer's outputs are the 1000 class scores for a different task like ImageNet), then treat the rest of the CNN as a fixed feature extractor for the pods dataset. CNN features are more generic in early layers and more

[1] https://cs231n.github.io/transfer-learning/tf

original-dataset-specific in later layers, as you've mentioned that it's hard to interpret single feature extracted by a neural network (whether it is efficient or useful for downstream classification tasks), hence we use the visual backpropogation method to interpret those features, we use Andromeda as a tool to enable users to form different projections based on the pods' Phenotypes, then the colormap generated by the visual backpropogation could illustrate whether those features could help distinguish data points or not by applying the learned weights. In future work, the resulting weighted feature models could then be used as classifiers for larger collections of images. We plan to extend these methods to more complex input scenarios, such as images of pods on live plants captured in the field with mobile phones.

⋮

Response To Reviewer #2

Overall Comments

> "The idea of 're-use' of a DNN is interesting and has application in phenomics. The authors did a proof-of-concept of DNN reuse with a study of edamame pods. The reuse concept could have been explored further, including how much is reusable and how much effort is saved vs. starting from scratch. The DNN is a basic pre-trained algorithm, but it is unclear how this model was trained as most information about the model is missing or assumed known by the reader. They show that the model adapts to changes in the input data, which is as expected. Most modeling choices are unexplained/unmotivated (e.g., why this DNN? why MDS and why reduction to 2D?). None of the classification results are provided (accuracy?) nor are the percentage of pods in each class in the data. Validation?"

Response

Thanks for the constructive comments. We discussed more about the modeling choices in our revised papers. We hope that you find the following response satisfactory.

Reviewer Comment

> "The reuse concept could have been explored further, including how much is reusable and how much effort is saved vs. starting from scratch."

Response

Thanks for the comments. How much is reusable and how much effort is saved are popular topics that computer vision researchers are doing nowadays for transfer learning. For this paper, we aim to provide investigative image sorting based on one fixed feature extractor. We aim to provide proofs that those features extracted by the pre-trained CNN model could capture human concept like "disease" or "number of seeds".

Reviewer Comment

> "The DNN is a basic pre-trainedalgorithm, but it is unclear how this model was trained as most information about the model is missing or assumed known by the reader."

Response

Thanks for the comments. We have add more reference talking about the pre-trained model that we chose for this paper.

Reviewer Comment

> "Most modeling choices are unexplained/unmotivated (e.g., why this DNN? why MDS and why reduction to 2D?)."

Response

We chose ResNet18 as the Neural Network of choice because of the consideration of model complexity as cited in the original paper. As supported by citations included in the original paper, ResNet18 is a powerful neural network model widely used in computer vision tasks. Furthermore, we can confidently claim that using ResNet18 over traditional convectional neural networks is certainly more desirable due to the maximum threshold depth in the traditional convolutional neural networks i.e. the training and testing error percentage rises if the threshold is crossed. For dimension reduction algorithms, MDS is easy to interpret and includes parameters that enable multiple forms of interaction. This is due to WMDS spatializations using distance between observations1 that reflect relative similarity; two points close to each other in a weighted MDS low-dimensional spatialization are considered more similar to each other in the high-dimensional space than are two points far from each other. We reduce the dimension into 2D to enable users implement human sorting, it is not able to do interactive tasks in multi-dimensional space. We cited more papers to support our modelling choices in the revised paper.

Reviewer Comment

> "None of the classification results are provided (accuracy?) nor are the percentage of pods in each class in the data. Validation?"

Response

Thanks for the comments, the validation for our approach is by observation level human perception. After the underlying model updating the feature weights, the users could see the new projection, the distance between data points indicate the similarities. Whether the data points could be grouped together by applying the learned weights is how we evaluate whether the CNN model capture the human concepts.

Response To Reviewer #3

Overall Comments

> "The approach appears novel and useful. A key weakness is that it is not clear how different methods are evaluated, or how this method performs relative to alternatives. Further, while a software system is described, there is no evidence that the software is available."

Response

We would like to thank you for you positive feedback. Your detailed comments have considerably helped with improving the clarity of the revised manuscript.

Reviewer Comment

> "A key weakness is that it is not clear how different methods are evaluated, or how this method performs relative to alternatives."

Response

Thank you for pointing out a valid concern. The reason there has not been any comparisons to alternative methods is because there is no other tool that has been implemented to allow investigative sorting of images. With that being said, we have specifically used this approach as it focuses on the user's cognitive spacial cues. In other words, such an intuitive approach would allow the user to draw a clear distinction between different groups projected on the Andromeda.

Reviewer Comment

> "Further, while a software system is described, there is no evidence that the software is available."

Response

Thank you for your feedback. We will be releasing an open-source computational notebook (Jupyter) through Binder that will contain the version of the Andromeda talked about in the manuscript.

Response To Reviewer #4

Overall Comments

> This is a proof-of-concept study of how a DNN trained for one classification task on visual images may be re-used for a similar classification task from the same images. The idea of getting input from users during the modeling process is very interesting and may save modeling time and resources. However, it is difficult to determine how classification accuracy changes with user input to the feature space, or how much user input is required for successful transfer of classification accuracy. We see image subspaces that are important to classification, but it is unknown how these compare pre- and post- user input. The amount of data shown is VERY small, particularly for CNNs, and no validation is provided.

Response

Thanks for the constructive comments. We hope that you find the following responses satisfactory.

Reviewer Comment

> "However, it is difficult to determine how classification accuracy changes with user input to the feature space, or how much user input is required for successful transfer of classification accuracy. "

Response

Preservation of accuracy and how much effort is saved are popular topics that computer vision researchers are exploring for transfer learning. For this paper, we aim to provide investigative image sorting based on one fixed feature extractor. We would like to provide proofs that those features extracted by the pre-trained CNN model could capture human concept like "disease" or "number of pods" through an interactive, cyclical process of human interaction with the data samples.

Reviewer Comment

> "We see image subspaces that are important to classification, but it is unknown how these compare pre- and post- user input."

Response

In a system that allows investigative image sorting, highlighting image sub-spaces that are important to classification allows human users to better understand the features that the underlying

model might use to represent certain distinguishing characteristics that they find in the data samples. This helps the human users to better understand if the underlying model produces a feature space which can represent characteristics of data samples that the users identify. For use cases in biology, the system might be used to determine if a model can represent a certain phenotype that researchers identify. Alternatively, researchers could interact with data samples in the system to investigate undiscovered phenotypes represented in the model's feature space. The paper has been updated with important image subspaces pre- and post- user input to demonstrate the value added through the investigative process.

Response To Reviewer #5

Overall Comments

> "The paper points to an interesting and potentially useful approach, but lacks sufficient information to adopt the approach and lacks quantitative results or comparison with related approaches."

Response

We would like to thank you for you positive feedback. Your detailed comments have considerably helped with improving the clarity of the revised manuscript.

Reviewer Comment

> "'The paper points to an interesting and potentially useful approach, but lacks sufficient information to adopt the approach and lacks quantitative results or comparison with related approaches."

Response

Thank you for pointing out a valid concern. We have specifically used this approach as it focuses on the user's cognitive spacial cues. In other words, such an intuitive approach would allow the user to draw a clear distinction between different groups projected on the Andromeda.

Response To Reviewer #6

Overall Comments

> "Need reference for claim that 'majority of parameters in the lower levels of the neural networks do not need to be retrained'. Relationship to transfer learning?"

Response

We would like to thank you for you positive feedback. Your detailed comments have considerably helped with improving the clarity of the revised manuscript.

Reviewer Comment

> "Need reference for claim that 'majority of parameters in the lower levels of the neural networks do not need to be retrained'. Relationship to transfer learning?"

Response

Thank you for your response.We indeed use the process of Transfer Learning in this project. The ResNet-18 is an 18 layer neural network trained on over a million images from the ImageNet database. Due to the high accuracy of the model, we decided to re-purpose a few higher layers of the neural network to make it relevant to our data set on hand. In fact, we re-purposed the neural network differently for the two data sets using transfer learning. In our research, our approach is to use multidimensional reduction to provide a 2D visualization for humans to provide interactive input that, through inverse dimensionality reduction, is imparted on the model through adjusted feature weights. Those features are features captured in the top-most layer(s) of the original model.