## Table of content

1. [Quick links](#_q313osrpf3o)
2. [Objective](#_aws342exyjov)
3. [Possible project title](#_1q105nlubuj7)
4. [Hypotheses](#_dgl4p3um9o88)
5. [Literature review](#_hgn154pkzg9j)
6. [Datasets](#_txd3etlkglee)
7. [Models](#_ipi407qf48df)
8. [Questions to TA](#_c0ghsul6h968)
9. [Draft of brainstormings](#_vhtaisw2h0o6)
10. [Summary of project formation and guidelines](#_kbubhkhkuizk)

## Quick links

* Github: <https://github.com/zahramor/NMA2022-dl-project>
* Template: <https://deeplearning.neuromatch.io/projects/Neuroscience/blurry_vision.html>
* Members:
  + Jahangir, [jahangir.esfandiari@esi-frankfurt.de](mailto:jahangir.esfandiari@esi-frankfurt.de)
  + Boris, [boris.sotomayor@esi-frankfurt.de](mailto:boris.sotomayor@esi-frankfurt.de)
  + Shaoyun, [shaoyun.yu@polyu.edu.hk](mailto:shaoyun.yu@polyu.edu.hk)
  + Zahra, [zahra.moradimanesh@gmail.com](mailto:zahra.moradimanesh@gmail.com)
  + Wei, [Wei.Shao@ruhr-uni-bochum.de](mailto:Wei.Shao@ruhr-uni-bochum.de)
  + Yang, [liny17@mails.tsinghua.edu.cn](mailto:liny17@mails.tsinghua.edu.cn)
  + Naresh, [nbrav@kth.se](mailto:nbrav@kth.se)

## Objective

How visual system in the brain solves noisy object classification? Imagine you lost your spectacle and the world around you is completely blurred out. As you stumble around, you see a small animal walk towards you. Can you figure out what it is? Probably yes, right? In this situation, or in foggy/night-time conditions, visual input is of poor quality; images are blurred, noisy, have low contrast and yet our brains manage to recognize it. Is it possible to model the process? Does previous experience help? How?

Already explored hypothesis in the template: If we pretrain a model on clear images and then train it on blurry images (the so-called expert learner), this surprisingly will not have much effect on its performance compared to the model that is only trained on blurry images (the so-called naive learner).

## Possible project title

1. Performance of naive learner vs. expert learner in convolutional neural networks in degraded images.
2. Modeling the effect of prior experience on noisy object classification in CNNs.

## 4. Hypotheses

1. [Non-trivial images]
   1. We hypothesize that the expert learner outperforms the naive learner in other kinds of images transformation: noisy (white noise), flipped, foggy/night-time conditions (low contrast), occluded, and specifically more complex (nonlinear) noises.
2. [Hyperparams]
   1. What is the impact of hyperparameters on the model’s performance? Maybe we don’t see much improvement in performance of expert learner compared with naive learners, because the important hyperparameters of the expert learner are fixed and not tuned. We hypothesize that the tuned expert learner can actually outperform the naive learner.
3. [Confident?]
   1. Is there any difference in the prediction probability (confidence) across the images? Is the expert learner less confident about the misclassified ones, or the non-trivial images? We hypothesize that the properties of the non-trivial images influence the prediction probability.
   2. [Something about RSA and other measures to evaluate internal representations]
4. [Finetune from the brain?]
   1. Maybe our brain uses different encoding for the non-trivial transformations. Does fine-tuning the CNNs with fMRI neuronal patterns (expert learner) help reach a better performance than the naive learner?
   2. Additionally, maybe if we fine tune the expert learner using fMRI bold responses the expert learner performs better than the naive learner. We hypothesize that the fine tuned expert learner using real bold signals can outperform the naive model. (Important note: to test this hypothesis we need an fMRI dataset with clear vs noisy images as stimuli.
5. [Increasing the number of pretraining images?]
6. The effect of experience (having seen clear images during pre-training, before training with noisy images) is helpful if only it is enough experience. We hypothesize that if we increase the number of clear images during pretarining the effect of experience will become evident. That is, the expert learner would outperform a naive learner.
7. [RNN addition to the CNN?]
   1. Once we trained the CNNs on the clear and blur images, the potential next step is to put the trained weight into the RNN network to find the sequential output by adding. Long Short Term Memory (LSTM) to be a good candidate to do this: <https://wiki.tum.de/display/lfdv/Recurrent+Neural+Networks+-+Combination+of+RNN+and+CNN>

## 5. Literature review

1. Abdelhack, Mohamed, and Yukiyasu Kamitani. "[Sharpening of hierarchical visual feature representations of blurred images.](https://www.eneuro.org/content/5/3/ENEURO.0443-17.2018)" *Eneuro* 5.3 (2018).
2. Tang, Hanlin, et al. "[Recurrent computations for visual pattern completion.](https://www.pnas.org/doi/abs/10.1073/pnas.1719397115)" *Proceedings of the National Academy of Sciences* 115.35 (2018): 8835-8840.
3. Grootswagers, Tijl, and Amanda K. Robinson. "[Overfitting the literature to one set of stimuli and data](https://www.frontiersin.org/articles/10.3389/fnhum.2021.682661/full)." *Frontiers in Human Neuroscience* (2021): 386.
4. Tovar, David A., et al. "[Getting the gist faster: Blurry images enhance the early temporal similarity between neural signals and convolutional neural networks](https://www.biorxiv.org/content/10.1101/2021.08.22.451834v1.full.pdf)." *bioRxiv* (2021).
5. ​Shi, Junxing, et al. "[Deep recurrent neural network reveals a hierarchy of process memory during dynamic natural vision.](https://onlinelibrary.wiley.com/doi/epdf/10.1002/hbm.24006)" *Human brain mapping* 39.5 (2018): 2269-2282.

## Datasets

1. The Asirra dataset: <https://www.kaggle.com/c/dogs-vs-cats>
2. Preproc datas: <https://openneuro.org/datasets/ds003661/versions/1.0.0>

## Models

1. Alexnet
2. Long Short Term Memory (LSTM)

## 8. Questions to TA

## 9. Draft of brainstormings:

## W1D4: Feel free to add new ideas/hypotheses, also papers, dataset, repo, etc.

* [Vision with Lost Glasses: Modeling how the brain deals with noisy input](https://deeplearning.neuromatch.io/projects/Neuroscience/blurry_vision.html)
* Human vision under foggy/night-vision circumstances are very robust. How do we use CNN for modeling this? A CNN is trained on only blurry images and compared with one trained on clean images and tested on blurry images for cat/dog classification.
* Asirra dataset containing images of cats and dogs is used. No additional preprocessing done.
* Additional preprocessing on data?
* Maybe we come up with different distortions, transformations on data and tests? Data augmentation?
* Adversarial examples are unique to CNNs (where changing a few pixels by adding noise leads to misclassification). How is this not the case for human vision? What are the internal representations for these examples (RSA, correlations, more)?
* We can do RSA (Representation Similarity Analysis) on CNNs. Schrimpf, Martin, et al. "Brain-score: Which artificial neural network for object recognition is most brain-like?." BioRxiv (2020): 407007.
* Adam optimizer is used, AdamW is a more advanced version that we can test. Also check the effect of other recent optimizers.
* Tune the learning rate, it’s fixed now (lr=3e-4).
* Tune batch\_size as well (batch\_size = 128).
* CE is used as a loss function, analyzing the impact of other loss functions.
* Review the literature to see if there are more recent models for visual cortex rather than [Schrimpf et al. 2020](https://www.biorxiv.org/content/10.1101/407007v2.full).
* What about other sources of noise, like Gaussion noise. Are there no meaningful differences between Naive and expert models?
* Limits to the RSA with regard to CNNs (<https://www.nature.com/articles/s41467-021-22244-7>)
* We can add many RNN layers within the CNN to detect the cortex response to a time-series task as inspired by this paper <https://onlinelibrary.wiley.com/doi/epdf/10.1002/hbm.24006>

* [Moving beyond Labels: Finetuning CNNs on BOLD response](https://deeplearning.neuromatch.io/projects/Neuroscience/finetuning_fmri.html)
* They computed (condensed) distance matrix of the 120 test images’ representations (dnet\_\* are based on activations of artificial neurons in a certain CNN hidden layer, dobs\_\* are based on BOLD voxel activations). They used the correlation between dnet\_\* and dobs\_\* as a measure of how good CNN layer activations can predict neural data (how similar are the distance matrices computed from BOLD signals and a certain CNN layer’s activations). A small step: try different Representation Similarity metrics?
* Which is the accuracy improvement of finetuning CNNs? Is this improvement dependent on how the input was blurred (e.g., gaussian kernels, add noise, apply other transformations)?
* [Focus on what matters: inferring low-dimensional dynamics from neural recordings](https://deeplearning.neuromatch.io/projects/Neuroscience/neuro_seq_to_seq.html)
* How good are RNN predicting these latent variables? If they are good, can RNN predict neural sequences during visual stimuli presentations (e.g., drifting gratings, natural scenes, movies). Is this prediction better for more simplistic visual stimuli (i.e, static gratings vs drifting gratings, natural scenes vs movies?
* (Combining Neuropixels data from Steinmetz et al): which are the latent variables of neural population responses during decision making? Are these latent variables consistent across brain areas? If not, can we define the contribution/relevance of each brain area based on these variables?
* We can introduce latent variables in the data generation that creates different LDS (like task0, task1, .. ) and ask if RNN can infer the latent task
* THis seems relevant: Mante, Valerio, et al. "[Context-dependent computation by recurrent dynamics in prefrontal cortex.](https://www.nature.com/articles/nature12742)" nature 503.7474 (2013): 78-84.
* Use [NMA task-fMRI data](https://colab.research.google.com/github/NeuromatchAcademy/course-content/blob/main/projects/fMRI/load_hcp_task_with_behaviour.ipynb#scrollTo=hC-v0oGNWcs_), which is clean and ready, to compare latent variables between for example the language task for math vs. the language task for story. The question would be is there any differences between the two tasks/conditions in terms of the latent variables? ([Modeling the Dynamics of Human Brain Activity with Recurrent Neural Networks, 2017](https://www.frontiersin.org/articles/10.3389/fncom.2017.00007/full), which also uses fMRI data)
* Neuropixels data (Steinmetz)
  + <https://www.nature.com/articles/s41586-019-1787-x>

## 

## W1D5: Questions and Hypothesis for Project “Vision with lost glasses”

**Boris’ ideas:**

1. How can this expert learner outperform the naive learner?
   1. Tuning hyperparameters
   2. Data Augmentation
   3. Only one type of noise was tested: Local noise? Different filters? non-linear combination of noise? Is the complexity of the blur or noise

H: Expert learner outperforms the naive one for inputs when blur/noise is more complex (linear vs nonlinear noise).

1. Is there any difference between the prediction probability of different images? If so, which are the properties of such images (Complexity)?
2. Internal representation of expert learners are more distinguishable than naive learner's one.
3. By training our CNNs with fMRI neuronal patterns, the expert learner will reach a better performance than the naive one

**Jahan’s ideas:**

1. We need some neurophysiological explanations: what happens when we see blurry or noisy images. With regard to different transformations, we might face different encoding in the brain. Based on the fMRI dataset that we will find, this could change.
2. CNNs trained or not trained on transformed images
3. Comparing expert and non-expert in CNNs

**Wei’s ideas:**

1. the difficult part is that we need to find a corresponding BOLD signals/voxel instead of image pixels as the input data? We need an experiment with two kinds of images that have a hierarchy structure. or :
2. We can implement the expert model(AlexNet ) into two different images (e.g. objects and faces) and find the BOLD signal pattern.(accuracy)
3. We can add an RNN layer to detect the temporal changes of BOLD signs when clear images and blur images are viewed. Would the pattern be different? Model comparison, by adding RNN layers, can we erase the gap of performance between trained CNNs and untrained CNNs?
4. And the can extrac the intermedia features from CNNs and RNNs both to answer the questions what factors determine the image viewing.

**Zahra’s ideas:**

1. Maybe we don’t see much improvement because the important hyperparameters of the expert learner are fixed and not tuned. We hypothesize that the tuned expert learner can actually outperform the naive learner.
2. The comparison between the naive and expert learner’s performance on only blurry image classification can be extended to other sources of noise. We hypothesize that the expert learner outperforms the naive learner in other kinds of noisy images, including flipped images, foggy/night-time conditions, etc.
3. Additionally, maybe if we fine tune the expert learner using fMRI bold responses the expert learner performs better than the naive learner. We hypothesize that the fine tuned expert learner using real bold signals can outperform the naive model. (Important note: to test this hypothesis we need an fMRI dataset with clear vs noisy images as stimuli.)
4. The effect of experience (having seen clear images during pre-training, before training with noisy images) is helpful if only it is enough experience. We hypothesize that if we increase the number of clear images during pretarining the effect of experience will become evident. That is, the expert learner would outperform a naive learner. (Note: to test this we need an extra animal dataset. I know one in Kaggle but it’s not labeled for cats and dogs, and we have to generate labels for it using a basic classifier if needed.)

**Naresh’s ideas:**

1. Humans robustly perceive and classify under various object transformations, distortions, occlusions? How far are CNNs robost to this and when and how do they fail? Maybe add white noise (of varying degree), occlusions (black boxes around the images), adversarial noise, blurring, rotations on the test set and look at the performance.
   1. What are the internal representations for these examples (RSA, correlations, maybe more)?
   2. Can training directly on these hard data samples improve model performance? (expert learner)
   3. One case where normal feedforward CNNs fail (but CNN+RNN can be better): [Tang, Hanlin, et al. "Recurrent computations for visual pattern completion." *Proceedings of the National Academy of Sciences* 115.35 (2018): 8835-8840](https://www.pnas.org/doi/abs/10.1073/pnas.1719397115). We need not have to implement RNNs, this is just to know the limitations of feedforward CNNs.

**Shaoyun’s ideas:**

1. Expert learner model’s intermediate layers’ activations may correlate better with BOLD signals (even without fine tuning on BOLD?) than the naive model does. Use similar methods used in the “Moving beyond Labels: Finetuning CNNs on BOLD response” example.

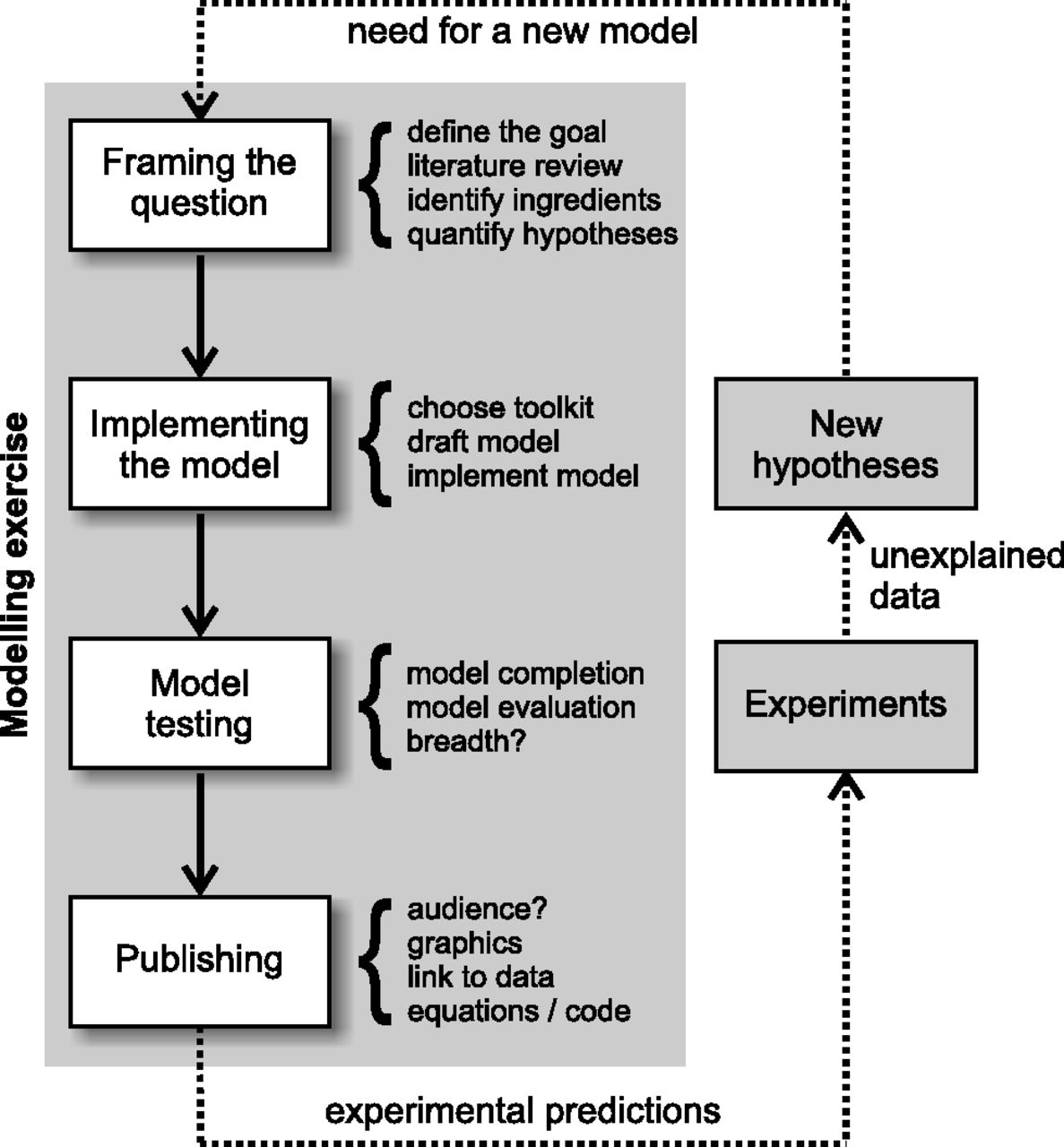
**Yang’s idea:**

Nothing to add, but a naive idea:

If we can find some datasets of fMRI or functional imaging of monkeys or mice watching images, could we restore the blurry images they really saw? (They have poor eyesight than human)

## 10. Short summary on project formation and guidelines

* series of questions and answers that gradually refine your hypotheses
* W1: look for an appropriate dataset to answer your question, and try to process that dataset into a format that is good for modeling
* W2: experiment with different types of deep neural networks to try to answer your question
* by W3D2:write a short abstract about the project, which may or may not include results, but it should at least include a testable hypothesis. For the rest of the week, focus on getting evidence for/against your hypothesis.
* W3D5 (last day of Project): l meet with other groups and tell them the story of your project, a low-key presentation that may include some of the plots you made along the way, but not meant as a real research presentation with high “production values”. See some of the [examples](https://compneuro.neuromatch.io/projects/docs/project_2020_highlights.html) from the comp-neuro course



Project Template

W1D4:

* Don’t start with the dataset: It may seem easier to start with a dataset + your favorite neural network, but that makes it very difficult to find a good question that matches that dataset. When you start with the question, you guarantee that you are personally invested in the question, and that you can have a unique angle or perspective that others may have not thought about yet.
* Finding a good dataset
* Finding the ingredients:
  + What parameters / hyperparameters / variables are needed?
    - Constants?
    - Do they change over space, time, conditions…?
    - What details can be omitted?
    - Constraints, initial conditions?
    - Model inputs / outputs?
  + Variables needed to describe the process to be modelled?
    - Brainstorming!
    - What can be observed / measured? latent variables?
    - Where do these variables come from?
    - Do any abstract concepts need to be instantiated as variables?
      * E.g., value, utility, uncertainty, cost, salience, goals, strategy, plant, dynamics
      * Instantiate them so that they relate to potential measurements!
* Linking the inputs and outputs
* Finding the hypothesis and formulate it mathematically:
  + You think about the hypotheses in words by relating ingredients identified in Step 3
    - What is the model mechanism expected to do?
    - How are different parameters expected to influence model results?
  + You then express these hypotheses in mathematical language by giving the ingredients identified in Step 3 specific variable names.

# 