METHODS

PROJECT OUTLINE

For each applicable step in the project cycle provide the (1) actions you will take (2) the justification for that action (3) the action you will take if the initial action fails.

We are looking only for action to be included in your final publication.

Example Action: Annotate 1000 frames with bounding boxes for male and female fish

Not an Action: Read literature about computer vision and fish.

Data Processing

Action	**Justification**	**Plan if Action Fails**	**Status**
Clean/transform mice Skull dataset to conform to the dataset expected by the pycpd algorithm	Pycpd looks to take a simple 3d point set, whereas the dataset provided is more detailed, which we may have to clean up	**n/a**	Done - looking to optimize data cleanup process.
Perform PCA on mice skull dataset to extract main features of the image set. We will use SciKit learn for this task, as it has a useful library for PCA.	In order to improve the coherent point drift algorithm and give it a biological basis, we have to use the PCA's to get the principal modes of variation and use this info for a basis to write a new kernel based on set input points.	Attempt to perform the same PCA with different library, or perform another shape modeling procedure (EFA maybe)	PCA is up and running, we have a reliable method that reports which features are taken, and uses them in a kernel
Analyze and keep enough of these modes of variation to explain 95% of variance	We need enough of the modes to explain 95% of the variation in the mean shapes to effectively replace	If this fails, then use EFA to do the same or similar process. Otherwise switch methods to SVD, where the points are	Done as part of the build_ssm method

	the current Gaussian Kernel.	not centered and allows for more sparse matrices.	

Algorithm Development

Action	**Justification**	**Plan if Action Fails**	**Status**
Inspect and test extant code in the Coherent Point Drift Space, focusing mostly on PyCPD, as this is the algorithm we will be editing/optimizing. We will test both the sample dataset provided by the library, and the dataset provided by Dr. Porto after cleaning.	The basis for our research is this code, which implements the coherent point drift algorithm	**n/a**	Inspected and the cpd process was replicated with the given dataset
Edit the extant PyCPD code, which replicates coherent point drift in Python, to accept a statistical model of a biologically valid specimen to serve as a basis for point set transformation	The motivation behind this is the primary goal for the project, if we can base the non-rigid areas of transformation on actual biological models, we should be more accurate and compute results faster for known datasets.	Attempt the same, but for bayesian CPD, which uses the same framework, and could lead to better results.	Finished a preliminary pca_registration method with the kernel replacement, but needs testing and optimization via hyperparameter tuning.

Test against extant pycpd and bcpd (Bayesian CPD) method for accuracy (RMSE) and compute time to see if there is any marked difference between the results for	This will give me a better idea of how the algorithm works, and any ways that I may be able to tune it to work with our specific datasets.	Use Hirose's original C code to test model if not feasible	Not started

Modal Selection/ Model Training

Action	**Justification**	**Plan if Action Fails**	**Status**
The model here is selected for us, which is the Python implementation of the coherent point drift algorithm.	Given that it surpasses most other geometric transformation models, this is expected to be the state of the art	There are a few models we can test, another that is a Bayesian Model, and another that is the same implementation in C.	Done

Train model on given labeled dataset, since both tasks are supervised learning approaches to solving this particular problem.	Supervised learning models must be trained	If not possible to train models on given data, tune models and edit them to be able to interpret some kind of success with given data.	Testing with given dataset was successfully completed
We must select a way to perform do principle component analysis, which I suspect we will be using SciKitLearn to do. These are CT scans of 50-100 mouse skulls, which are to be processed	PCA must be performed on the dataset to extract most pertinent components to base the new kernel instead of the Gaussian default	Find a different way to extract dominant correlation patterns which explain 95% of variation (required), SVD is a current option	The PCA process is finished

Validation of method

Action	**Justification**	**Plan if Action Fails**	**Status**
Use the in-sample real biological dataset to validate the model. This will involve comparing the augmented CPD with the vanilla one provided in pycpd, and the bayesian method.	Models must be validated	Restart training stage and tune new kernel model until promising results	Done

Evaluate precision and accuracy via loss measurements using negative log likelihood and RSS	There is a necessary metric against which to measure loss against to decide how successful the model is	Use other metrics to quantify loss	Started on a testing suite that takes RMSE's of models and compares them one against one
Once initial RMSE tests are performed, we can potentially move onto testing compute time between the different methods.	The hypothesis is that the model will take less time to train while adding computational complexity, since points will be closer to their realistic mark and will have to transform less, which means less EM iterations in the GMM step	If failed, attempt to discuss why and reformat the algorithm to get it to a faster working speed.	Compute time is part of the testing suite, but not explicitly tested yet

User-interface/ User-experience development

Action	**Justification**	**Plan if Action Fails**	**Status**
Work on providing data visualization examples using Slicer, to illustrate success/improvemen t of the model	May help data be visualized clearer by interested parties.	n/a	Not started
Publish a demo video or gif to the website for our research	Visual representation is necessary in a computer vision project, and can serve to demonstrate improved accuracy	Take a video and screen record differences to post online	Not started