

# Project Proposal: Category-Agnostic Pose Estimation

- **Students IDs, names, and e-mail addresses.**
  - NetID: nk699, Niki Karanikola, [nk699@cornell.edu](mailto:nk699@cornell.edu)
  - NetID: tc796, Theodoros Chronopoulos, [tc796@cornell.edu](mailto:tc796@cornell.edu)
  - NetID: pr484, Pavlos Rousoglou, [pr484@cornell.edu](mailto:pr484@cornell.edu)
- **What paper/papers did you choose?**
  - - CAPE paper (first work): <https://arxiv.org/abs/2207.10387>
  - - CapeX: <https://arxiv.org/abs/2406.00384>
  - - Raster2Seq: Hao Phung paper
- **What is the problem that you will be investigating? Why is it interesting?**
  - We're investigating category-agnostic pose estimation. For example, given a single image and a simple pose graph (a short list of keypoints and how they connect) we want to predict 2D keypoints for objects from unseen categories, without relying on labeled "support" images and category labels. This is interesting because classic pose models are usually tied to one category (e.g., humans) and need retraining and new annotations to handle anything new. A category-agnostic method can generalize "out of the box", which lowers labeling cost. If category-agnostic pose estimation works well, it could power many applications (e.g. robotics and warehouses, sports and motion analysis, safety and surveillance, fall-risk or hazard detection) by letting us add new object types simply by defining their pose graphs instead of collecting thousands of labels.
- **What data will you use?**
  - We will use the MP-100 dataset for training and testing our model, which is a 2D pose dataset of 100 object categories containing over 18K instances, consisting of various subjects such as animals, vehicles, furniture, and clothes.
- **What method/model will you use? Are there any existing implementations? If so, in which libraries are they written?**
  - The methodology is inspired by the CAPE problem. We will adapt the Raster2Seq framework using Hao Phung's codebase to perform category pose estimation using only a query image and a pose graph represented as a keypoint sequence, eliminating the need for annotated support images.
- **How will you evaluate the results?**
  - We will evaluate the results using PCK (Probability of Correct Keypoint), which measures how often a predicted keypoint lands *close enough* to the ground-truth keypoint. We will use PCK as our primary metric because it is the standard evaluation measure adopted by virtually all prior category-agnostic pose estimation papers, enabling direct and fair comparison with existing work. We will compare our methodology's results with those from ProtoNet, MAML, CapeX, POMNet, and CapeFormer under the 1-shot setting.
- **What will be the innovation in your project? Though this may change while working on the project, it is important to have a goal from the beginning**
  - The key innovation from the CapeX lies in the pose-graph representation which we will introduce: while CapeX employs textual descriptions of keypoints, our method directly utilizes 2D coordinate sequences as the support data for pose estimation.