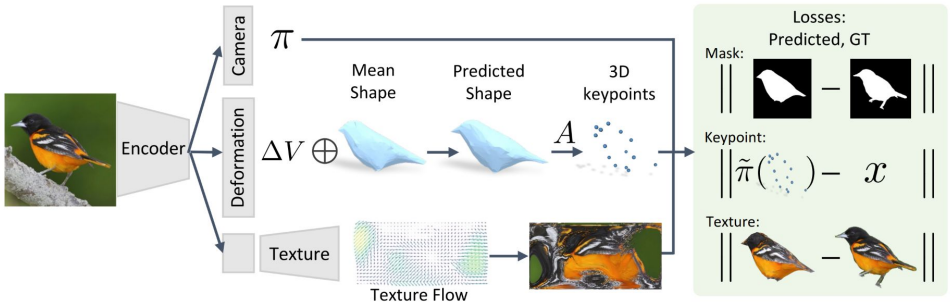
Learning Category-Specific Mesh Reconstruction from Image Collections

Group 4: Rishikesh Jadhav, Nishant Pandey(scribe), Shantanu Parab, Archana Swaminathan, Andrew Zhong

Approach

The paper presents a learning framework for recovering the 3D shape, camera, and texture of an object from a single image.

- 3D Shape and Texture Prediction
- Shared Mesh Structure with UV Mapping
- Regularization and Keypoint Association



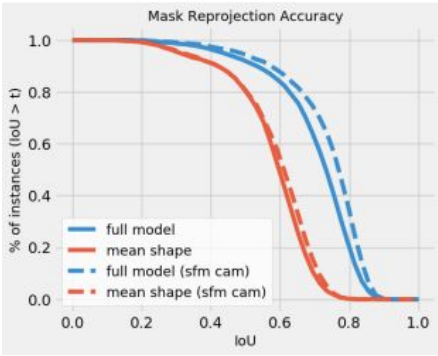
Experimental Setup

1. Dataset Selection: CUB-200-2011 dataset, which comprises 6000 training and test images spanning 200 species of birds.
2. Filtering: Nearly 300 images were filtered out from the dataset. The criterion for this exclusion was images where the visible number of keypoints was six or fewer(truncated close shots).
3. Data Splitting: They divide the test set in half to create a validation set, which they use for hyper-parameter tuning.
4. Network architecture: The encoder part of this network is built upon an ImageNet pretrained ResNet-18, followed by a convolutional layer. Which is sent to two fully-connected layers. On top of this there are linear layers to get the camera pose and deformation. The texture flow component consists of 5 upconvolution layers where the final output is passed through a tanh function.

Results

- Accurate but struggles to capture rare poses.
- Has issues with scenarios where the bird has open wings, etc.
- The full predicted shape slightly outperforms using only the mean shape.
- In comparisons with other approaches using the PASCAL 3D+ dataset, their approach achieves similar or better IoU results while also providing additional outputs such as texture.
- Is pretty accurate the texture transfer on the body of one bird to another is consistent.

Texture transfer results



Pascal 3D+ results

Strengths

- Able to learn from annotated image collections without needing any 3D ground truth.
- It is able to generate textures onto the output. Other approaches don’t.
- Was able to generate good results qualitatively and quantitatively for a challenging object category not commonly addressed in other approaches.
- It was able to perform similarly to other methods in rigid classes in PASCAL 3d+.

Weaknesses

- Inability to predict rare poses.
- Does not incorporate asymmetric articulation.
- The lack of a 3D ground truth for benchmarking on the CUB dataset hinders the true measure of the performance of the framework on unique object categories.
- The framework is sensitive to amount of training data for specific categories and struggles if a shiny region is present in source image.