USING DIFFUSION FEATURES FOR OBJECT POSE ESTIMATION

Carnegie Mellon University

AMAN CHULAWALA, BEVERLEY-CLAIRE OKOGWU

TEAM 39

SCOPE OF OUR WORK

1. **Pose Estimation:** Develop a robust method for estimating the 6D pose of objects from images, a crucial task in 3D scene understanding.
2. **Generalization to Unseen Objects:** Address the issue of limited generalizability of image features in current methods, which leads to a significant performance drop when dealing with unseen objects.
3. **Diffusion Model Features:** Conduct an analysis on the features of diffusion models, such as Stable Diffusion.
4. **Performance Evaluation:** Evaluate the proposed method on LINEMOD dataset and compare its performance with state-of-the-art methods.

METHODOLOGY

Dataset Preprocessing

Calculate GT poses
Rendering Templates
Crop the Template
Compute neighbors with GT poses
Creating Dataset Splits

↓

Training of the Model

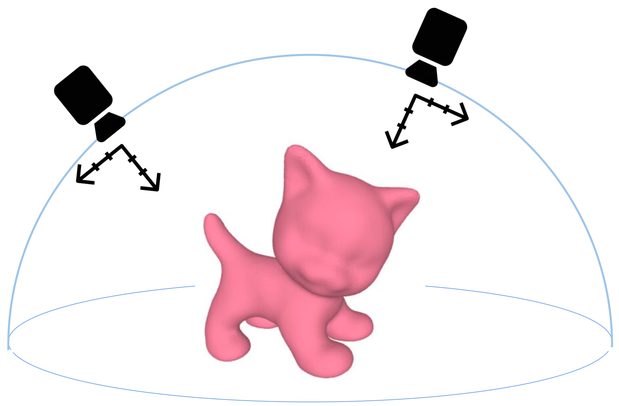
Diffusion Feature Extractor
Aggregation Network
Training Pipeline

↓

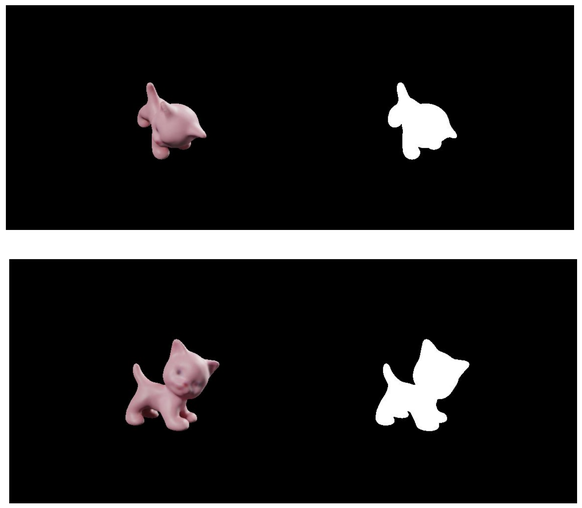
Testing of the Model

Testing Metrics
Visualization

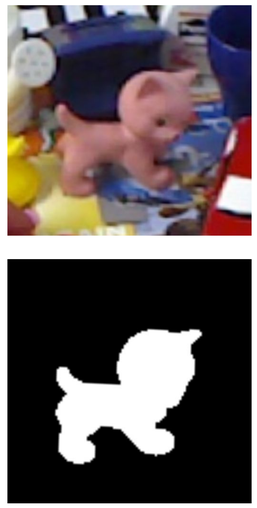
Cat model from LINEMOD



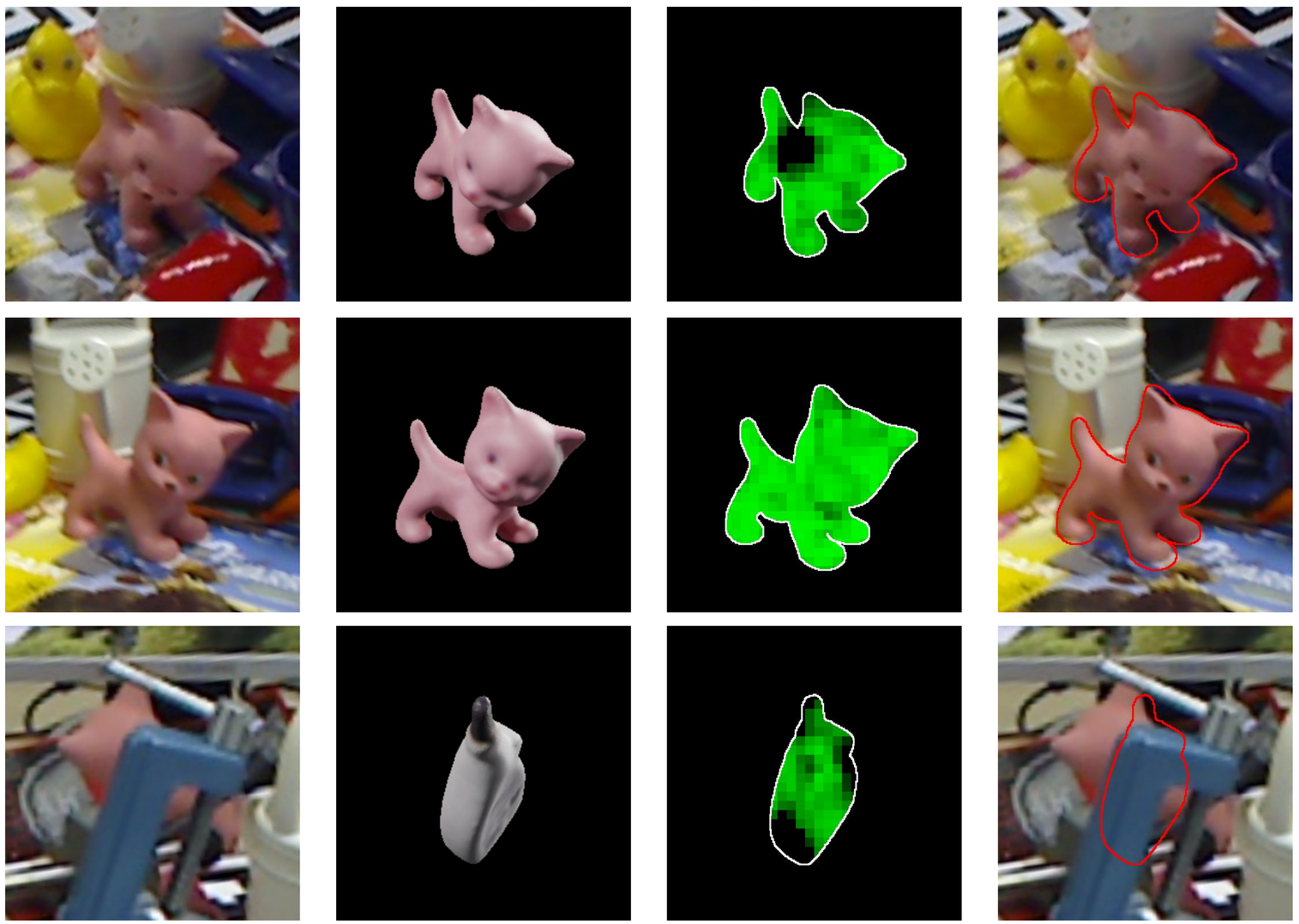
Half Sphere used to Sample Views



Rendered GT Templates



Queried Pairs



CURRENT RESULTS

1. **Dataset preprocessing finished:** We have finished setting up the dataset. Creating the rendered ground truth templates took up the majority of time. We have rendered the ground truth templates and cropped them for training. Dataloader have been set up.
2. **Working Training pipeline:** Implemented Diffusion Feature Extractors and Aggregation Methods based on existing works. Set up a standard training pipeline using InfoNCE loss.
3. **Working testing pipeline:** Implemented a testing pipeline which makes use of the trained models to perform inference on the dataset.
4. **Model trained on random hyperparameters:** Trained a standard model with no hyperparameter optimisation, which provides sub-optimal results.

FUTURE WORK

1. **Optimize hyperparameters:** Find a good combination of hyperparams which result in a better performing model.
2. **Improve the model architecture:** The current model architecture references on previous works with notable deficiencies, which can be improved
3. **Compare against existing algorithm:** Compare pose results against existing method created which does not rely on learning based methods.