P2Net: Patch-match and Plane-regularization for Unsupervised Indoor Depth Estimation [Link]

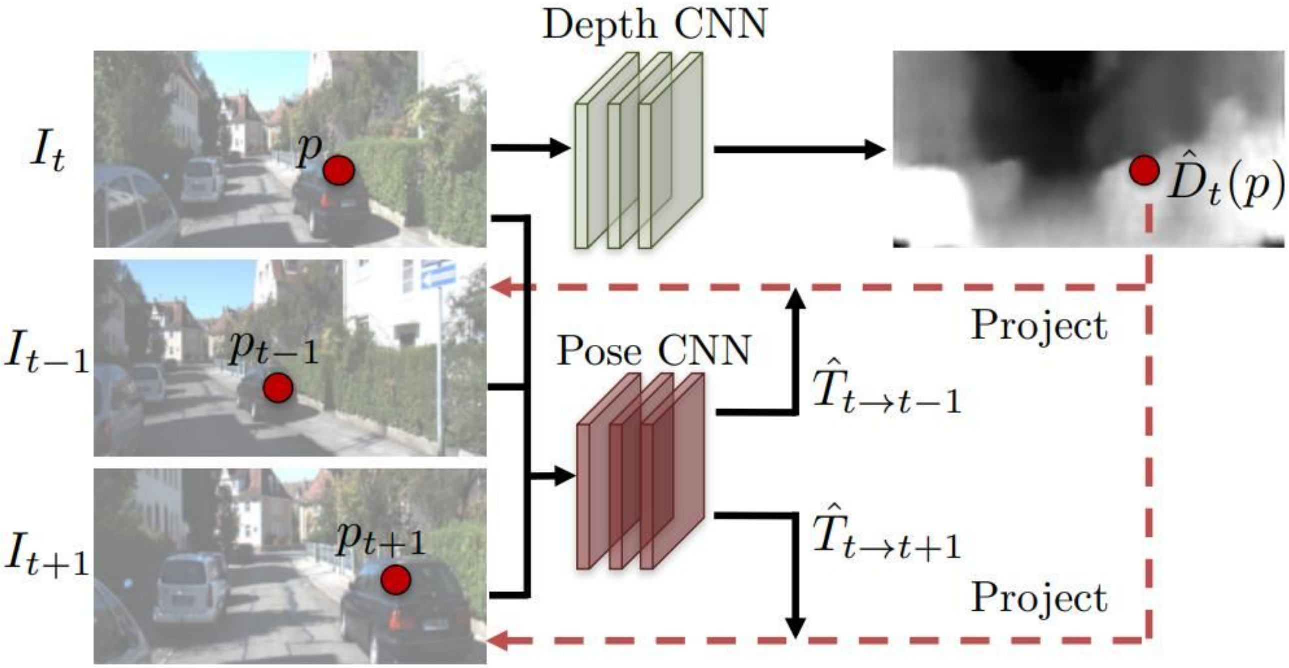
Shantanu Singh
Kaustab Pal
Kollepara Neeraj
Mohammad Abdul Mannan

Objective

The objective of the project is to learn depth estimation for indoor scenes using only monocular rgb images in an self-supervised manner similar to SFM.

Method overview

The method is based off of the SFMLearner technique:



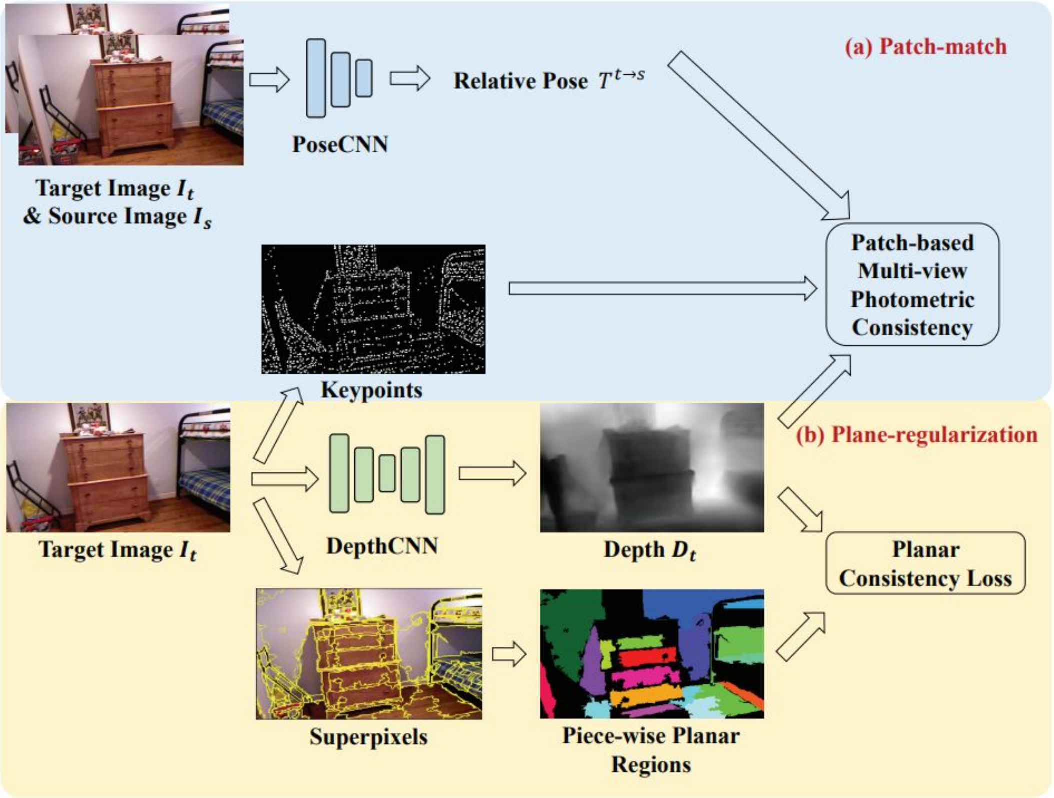
Method overview

- The method consist of using 2 neural networks that do depth and relative pose prediction given their respective inputs.
- The depth network takes only the view I_t as input, and outputs a per-pixel depth map D_t.
- The pose network takes both the view I_t and its neighbouring views I_{t-1} and I_{t+1} as input, and outputs the relative camera poses $T_{t\to t-1}$ and $T_{t\to t+1}$.
- The poses are then used to inverse warp the view I_{t-1} and I_{t+1} to reconstruct the view I_t, and the photometric reconstruction loss is used for training the networks.

Method overview

- Since indoor environments have a lot of textureless regions, the photometric reconstruction loss function will serve as a noisy signal. To reduce this, we simply use a point detector to extract keypoints and then use patches around these key points as valid regions for computing the loss function.
- Another technique that can be used for stabilizing the learning is to enforcing the planarity constraint on the textureless regions. For this, we first segment the image as superpixels and filter out those that aren’t large enough. Then using the predicted depth map, we try to estimate the plane parameters for points that belong to each superpixel and enforce this by adding a loss term that penalizes for discrepancy between the estimated planar depth and predicted depth.

Method overview



Goals

- For the first phase, we want to implement the SFMLearner architecture. This will serve as a baseline and help us gauge how the various techniques like using patches around keypoints, and enforcing planarity using superpixels improve the result.
- In the second phase, we'll implement the patches around keypoints technique and evaluate its performance on different hyperparameters like patch window size, number of keypoints etc.
- In the last phase, we'll implement the planarity constraint using the superpixels. Here we'll gauge how the quality of superpixels affects the depth estimation.

THANK YOU