1. Main ideas
   1. Perceiving the nature of the 3D scenes: consisting of **static rigid structures** and **dynamic object motions**
   2. Rigid Structure Constructor
      1. DepthNet
         1. Input single view to predict the depth
      2. PoseNet
         1. Input a sequence of views to predict relative poses from source views to a target view
   3. Non-Rigid motion localizer
2. Loss
3. Network architecture
4. Implementation details
   1. INPUT:
      1. Image/flow pyramid: scale
      2. Stack multiple source views
   2. OUTPUT
      1. Image/flow pyramid: rescale
   3. Image warping and its differential computation
5. Datasets
   1. Format
   2. Data Loader
6. Training
   1. Training examples: temporal continuous frames and known camera intrinsics
7. Inference
8. Evaluation
   1. Metrics
   2. Timing
9. Considerations outside the papers
   1. Why smooth in the reconstruction (Depth estimation)?
   2. If we want to reconstruct a complete 3D static rigid structure, but the input views contains occlusion, how to do that? 3D inpainting?
   3. 这样的net他其实并没有回环检测, 想象一下,如果在一个很crowded 的environments里mapping,可能这个robotics不能在一次经过某地的时候就将这个地方mapping完整,但是多次经过就会不断的complete 这个scene? 所有loop closure怎样在dnn中实现呢? 虽然有单独做loop closure的,但是如果将这些component全部用一个完整的dnn实现出来,又应该怎么做呢? 有可能吗? 有必要吗?
   4. Depth maps 的estimation并没有构建一个完整的3D scenes, 而只是一帧一帧图像的depth estimation,难道后面这个3D scenes的构建是trivial的? 不然为什么省略呢? 或者是问题的目标不同? 这真的不是一个SLAM问题啊
   5. $$L\_{rw}$$里面计算了SSIM,虽然是评价结构相似度的,但是这是否会受到moving objects的影响呢,显然是会的. 所以这个loss也并不能完整地刻画这个loss所要表达的意思, 显然的