Novel view synthesis is a challenging problem in computer vision and robotics. Different from the existing works, which need the reference images or 3D models of the scene to generate images under novel views, we propose a novel paradigm to this problem. That is, we synthesize the novel view from only a 6-DoF camera pose directly. Although this setting is the most straightforward way, there are few works addressing it. While, our experiments demonstrate that, with a concise CNN, we could get a meaningful parametric model that could reconstruct the correct scenery images only from the 6-DoF pose. To this end, we propose a two-stage learning strategy, which consists of two consecutive CNNs: GenNet and RefineNet. GenNet generates a coarse image from a camera pose. RefineNet is a generative adversarial network that refines the coarse image.

We propose a two stage framework consist of GenNet and RefineNet that connect in a sequence to synthesize an image from a camera pose. In the first stage, we use GenNet to get a coarse from pose P. In the second stage, we use RefineNet to refine the coarse image to get a fine detailed image.

- **GenNet**
  - Use two fully connect layer to expand pose alone channel dimension
  - Use U-Net structure, following pix2pix. The loss function includes L1 norm, Perceptual Loss and Adversarial Loss.

- **RefineNet**
  - We use U-Net structure, following pix2pix. The loss function includes L1 norm, Perceptual Loss and Adversarial Loss.

- **On Cambridge Landmark**
  - Examples of results on 7Scenes. For each scene, the first row contains synthesized images, and second row contains corresponding ground truth images.

- **Ablation Studies**
  - The effect of RefineNet. The first row is coarse image and second row is refined image (Left)
  - The effect of Perceptual Loss. PL could help to remove unrealistic artifact (top)

- **On 7-Scenes Dataset**
  - Examples of results on 7Scenes. For each scene, the first row contains synthesized images, and second row contains corresponding ground truth images.

- **Conclusion**
  - We propose a new problem configuration of NVS: take only the camera pose as input. A two-stage framework consist of two consecutive networks: GenNet and RefineNet is used. Experiments show promising results in generating visually pleasant images. There are also limitations: need to train network for each scene; distortion in results and limited generalization ability.