Q1.

Our experiments observed that the method using feature points is much better than the one using optical flow. We noticed some artifacts in the output for the latter approach. Also, the motion smoothing and stabilization are much better in the case of the feature points method. The quality of results may vary according to the parameters we use for smoothing. Larger values for variance or window size of the smoothing filter leads to introducing more artifacts in the videos, even if the motion stabilization is better.

Q2.

a) When data(Q2) (both object and camera moving) used in code(Q1) ,
Since we gave the manual threshold while applying the RANSAC , the features of the shaky object were not getting selected.

b) g
c) f

Q3.

a) The motion tends to be erratic in the videos in the Q3 folder. In some intervals in the video, the camera moves very fast due to which sudden changes in the scene occur. Hence, simple averaging filters like Gaussian and moving average filters with constant parameters are not able to reduce such spikes in the motion of the camera.

b) Adaptive smoothing techniques like Adaptive Gaussian filter may aid in mitigating such problems. In case of Adaptive Gaussian filter, the value of standard deviation (σ) varies (within a window W) with change in x:

$$\sigma_{adaptive} = \sqrt{\frac{1}{N-1}\sum_{(x,y)\in R}(f(x,y) - average)^2}$$

So, $\sigma_{adaptive}$ increases with increase in the speed of the camera. So the spikes in the motion can be more effectively flattened out (smoothed).

c) Some possible problems with amateur videos are:

1. **Moving/shaking camera**: Smoothing motion of feature points (using homography matrix) or activity for every pixel (dense optical flow)
2. **Moving elements in the background** like waves in a sea, rustling leaves, birds, etc.: Eliminating feature points for moving objects. Various methods can do this. One such process involves calculating optical flow for feature points/ all pixels and computing the histogram for the magnitude of these flow vectors. Then, the most common volumes, i.e., magnitude values with the highest no. of pixels, are considered the static background pixels as all the pixels corresponding to static objects move with almost the same velocity.

3. **Poor illumination** or **sudden changes in lighting** (like clouds obscuring the sun, flickering light, etc.)
4. **Poor exposure**
   Contrast stretching methods for each frame can help in case of problems mentioned in 4 and 5.
5. Poor focus in some scenes

Q4.

a) Videos inside output folder

b) We can use the combination of the multi-stage deep learning framework as a coarse-dense net (Eigen et al.). First off, the input video goes through the coarse net and then fine-network to get the more acceptable resolution video.

The coarse (initial) stabilizer calculates the rigid transformations among frames to alter a stabilization structure globally.

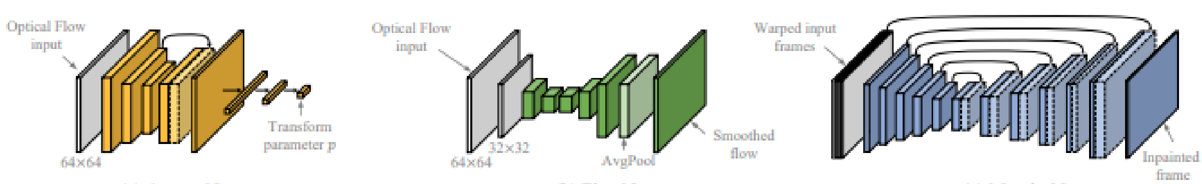
The fine (second) stabilizer adjusts the residual instability by estimating the spatially smoothed flow between frames.

Finally, the margin in painter fills in the frame borders to eliminate cropping.

The stabilized frames are produced by successive feed-forward passes of the coarse and fine networks, proven effective.

The whole framework can be used as the self-supervised framework,
And the consecutive frames can be considered as weak labels.
The appropriate loss function to train the network.

Reference [here]



(a) Coarse-Net (b) Fine-Net (c) Margin-Net