Mohit Kumar
SR No. 04-01-03-10-51-21-1-19825
MTech Artificial Intelligence

A2: Video Prediction using Deep Optical Flow

In this assignment we will predict a future video frame given the two past frames using deep optical flow. We will estimate the forward optical flow (flow from Frame n to Frame n+1) and assume linearity of motion to predict the future frame (Frame n+2) on corridor and sphere datasets. To estimate the optical flow we use pretrained FlowNet2 deep learning model and experiment by performing test time adaptation of the optical flow algorithm on a given pair of frames from the corridor and sphere datasets. We observe that the quality of the predicted frames is comparable to the original for both the sphere and corridor datasets.

I have reported mean square error and SSIM metric to compare the predicted and ground truth video frames.

Note: In all figures of predicted frames the left image is the predicted frame and the right image is the ground truth frame. Due to space constraints only limited figures have been included in the report. All the predicted frames and optical flow maps can be found in the zip file submitted along with this report.

- **Optical Flow prediction using pretrained FlowNet2 model** I have used the FlowNet2 model made publicly available by NVIDIA.
  Figures 1 and 2 show the predicted frame and predicted optical flow for the corridor dataset. We observe that the quality of the predicted frame is really good compared to that of the classical optical flow methods used in Assignment 1. We observe some error at the edges which can be attributed to pixels moving in or out of the frame.

  Figures 3 and 4 show the predicted frame and predicted optical flow for the sphere dataset. We observe that the quality of the predicted frame is really good and is marginally better than the classical optical flow methods used in Assignment 1, which were also of good quality. No visual error can be observed.

– *corridor dataset*

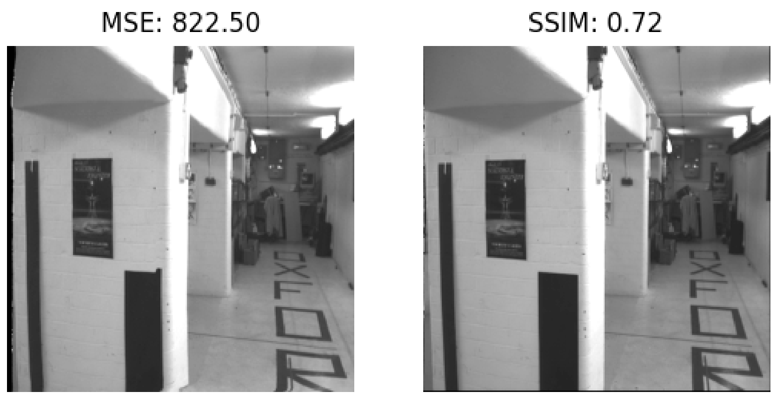

Figure 1: frame predicted by pretrained model

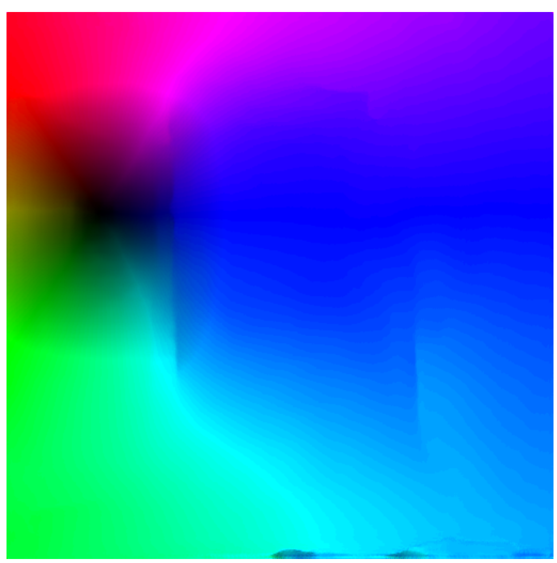

Figure 2: flow predicted by pretrained model

- *sphere dataset*

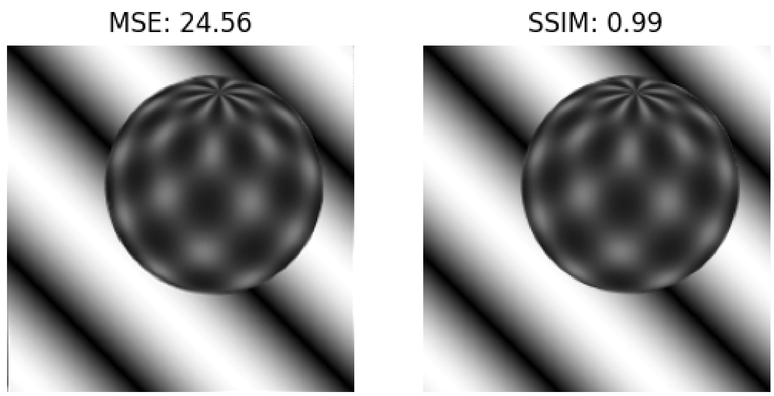

Figure 3: frame predicted by pretrained model

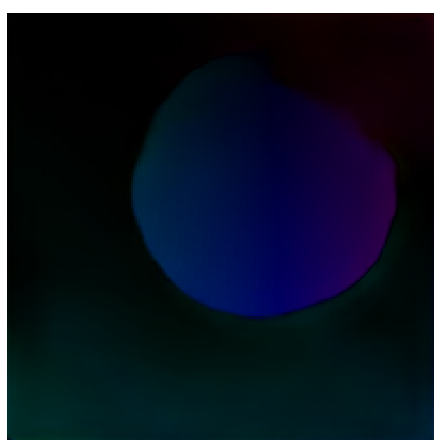

Figure 4: flow predicted by pretrained model

- **Test Time Adaptation of Optical Flow Algorithm** Now we discuss frame prediction performance after test time adaptation of the the pre-trained FlowNet2 model made pub-

licly available by NVIDIA.
We use an unsupervised loss function which is the sum of photometric loss and smoothness loss, as the loss function for the test time adaptation.
Figures 5 and 6 show the predicted frame and predicted optical flow for the corridor dataset with deep optical flow estimation after test time adaptation(for 10 iterations). We can observe that the quality of the predicted frames is really good compared to that of the classical optical flow algorithms. However, no significant visual quality improvement was observed over the pre-trained model results which can be attributed to the well trained model. No significant visual quality improvement was observed with respect to change in the number of iterations.

Figures 7 and 8 show the predicted frame and predicted optical flow for the sphere dataset with deep optical flow estimation after test time adaptation(for 10 iterations). Again, the quality of the predicted frames is really good and marginally exceeds the quality of the classical optical flow algorithms of Assignment 1(which itself was of good quality). Again, no significant visual quality improvement was observed over the pretrained model results, which can be attributed to the well trained model. No significant visual change in quality was observed with respect to change in the number of iterations. We use the unsupervised loss function proposed in [3], which is the sum of photometric loss and smoothness loss, as the loss function for the test time adaptation.

$$\mathcal{L}(\mathbf{u}, \mathbf{v}; I(x,y,t), I(x,y,t+1)) = l_{\text{photometric}}\left(\mathbf{u}, \mathbf{v}; I(x,y,t), I(x,y,t+1)\right) + l_{\text{smoothness}}\left(\mathbf{u}, \mathbf{v}\right)$$
$$l_{\text{photometric}}\left(\mathbf{u}, \mathbf{v}; I(x,y,t), I(x,y,t+1)\right) = \sum_{i,j} \rho\left(I(i,j,t) - I\left(i+u_{i,j}, j+v_{i,j}, t+1\right)\right)$$

$$l_{\text{smoothness}}\left(\mathbf{u}, \mathbf{v}\right) = \sum_{j}^{H} \sum_{i}^{W} \rho\left(u_{i,j} - u_{i+1,j}\right) + \rho\left(u_{i,j} - u_{i,j+1}\right) + \rho\left(v_{i,j} - v_{i+1,j}\right) + \rho\left(v_{i,j} - v_{i,j+1}\right)$$

We consider the penalty function to be Charbonnier function $\rho(x) = \left(x^2 + \epsilon^2\right)^{\alpha}$.

– *corridor dataset*

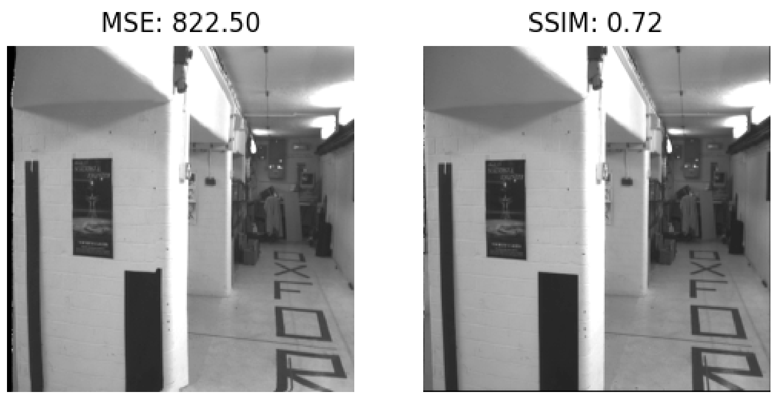

Figure 5: frame predicted after test time adaptation

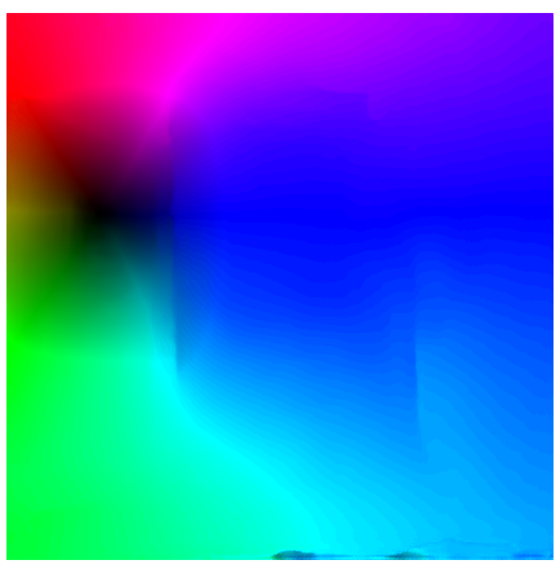

Figure 6: flow predicted after test time adaptation

– *sphere dataset*

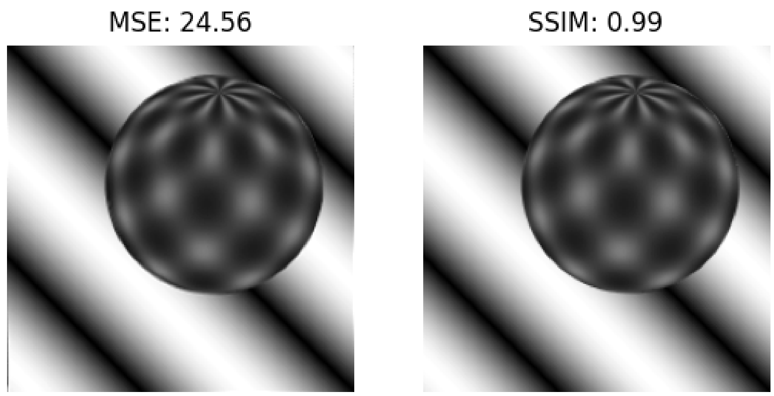

Figure 7: frame predicted after test time adaptation

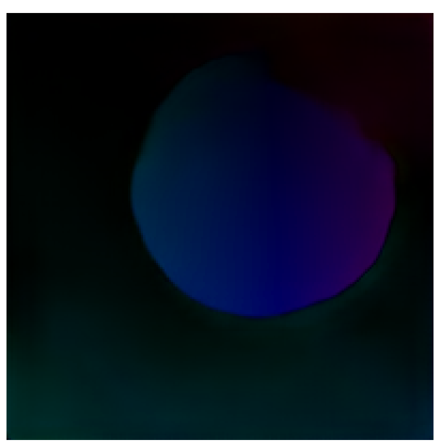

Figure 8: flow predicted after test time adaptation

Comparison between Classical Optical Flow Estimation and Deep Optical Flow Estimation

In the previous assignment, we observed that classical optical flow algorithms performed well on the sphere dataset and poorly on the corridor dataset. In this assignment, we observed that FlowNet2 which is a deep optical flow algorithm performed well on both the sphere and corridor datasets. There was a significant improvement in the visual quality of the predicted frame with optical flow prediction using FlowNet2 for the corridor dataset. However, there was no significant visual quality improvement with the sphere dataset by using FlowNet2 to estimate the optical flow. We can therefore conclude that deep optical flow estimation such as the FlowNet2 model generally perform better than classical optical flow algorithms.

References [1] Eddy Ilg et al. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks. 2016. arXiv: 1612.01925

[2] Fitsum Reda et al. flownet2-pytorch: Pytorch implementation of FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks. https://github.com/NVIDIA/flownet2-pytorch. 2017.

[3] Jason J. Yu, Adam W. Harley, and Konstantinos G. Derpanis. Back to Basics: Unsupervised Learning of Optical Flow via Brightness Constancy and Motion Smoothness. 2016. arXiv: 1608.05842