

Frame Booster

Kamil Barszczak

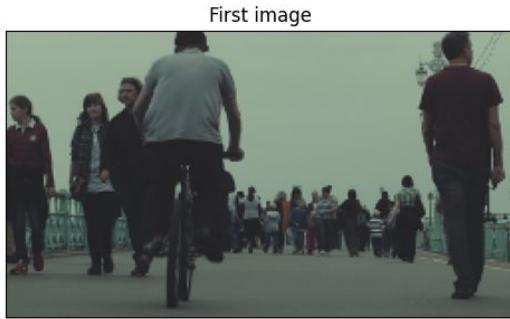


Frame interpolation

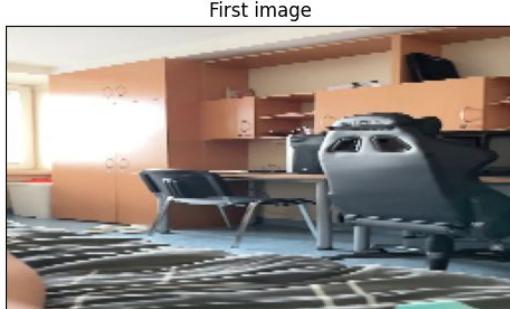


The dataset used during the training

- Vimeo90K - triplet (~ 51000 samples, 448x256x3, 35 GB)

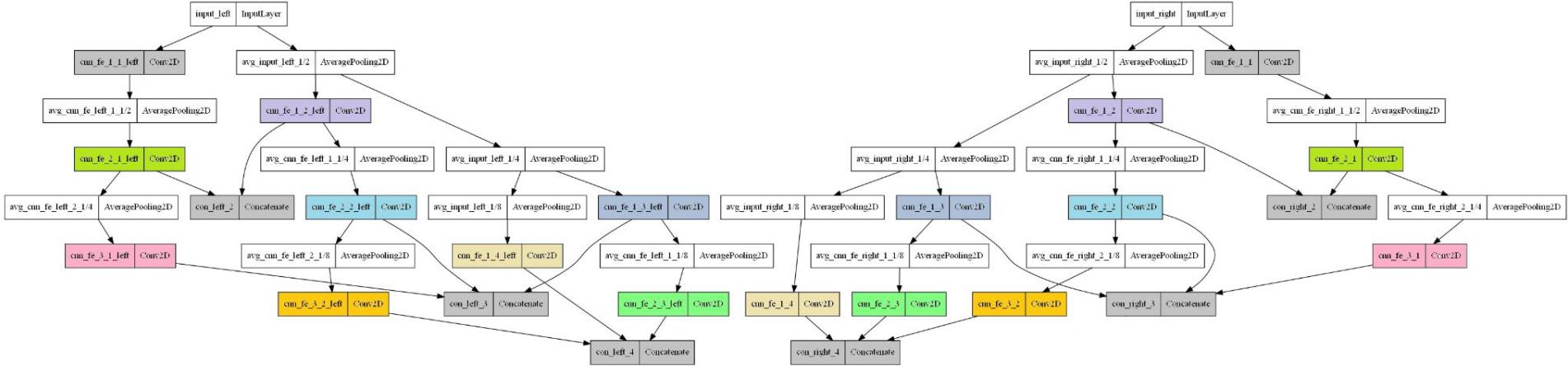


- Custom data (~ 1000 samples, 256x144x3, 1.3 GB)



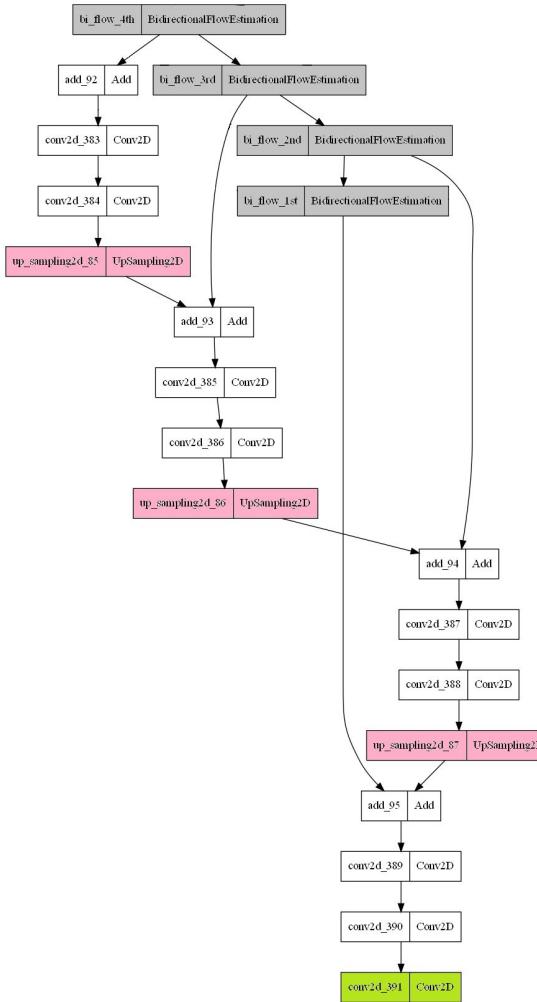


FBNet - encoder

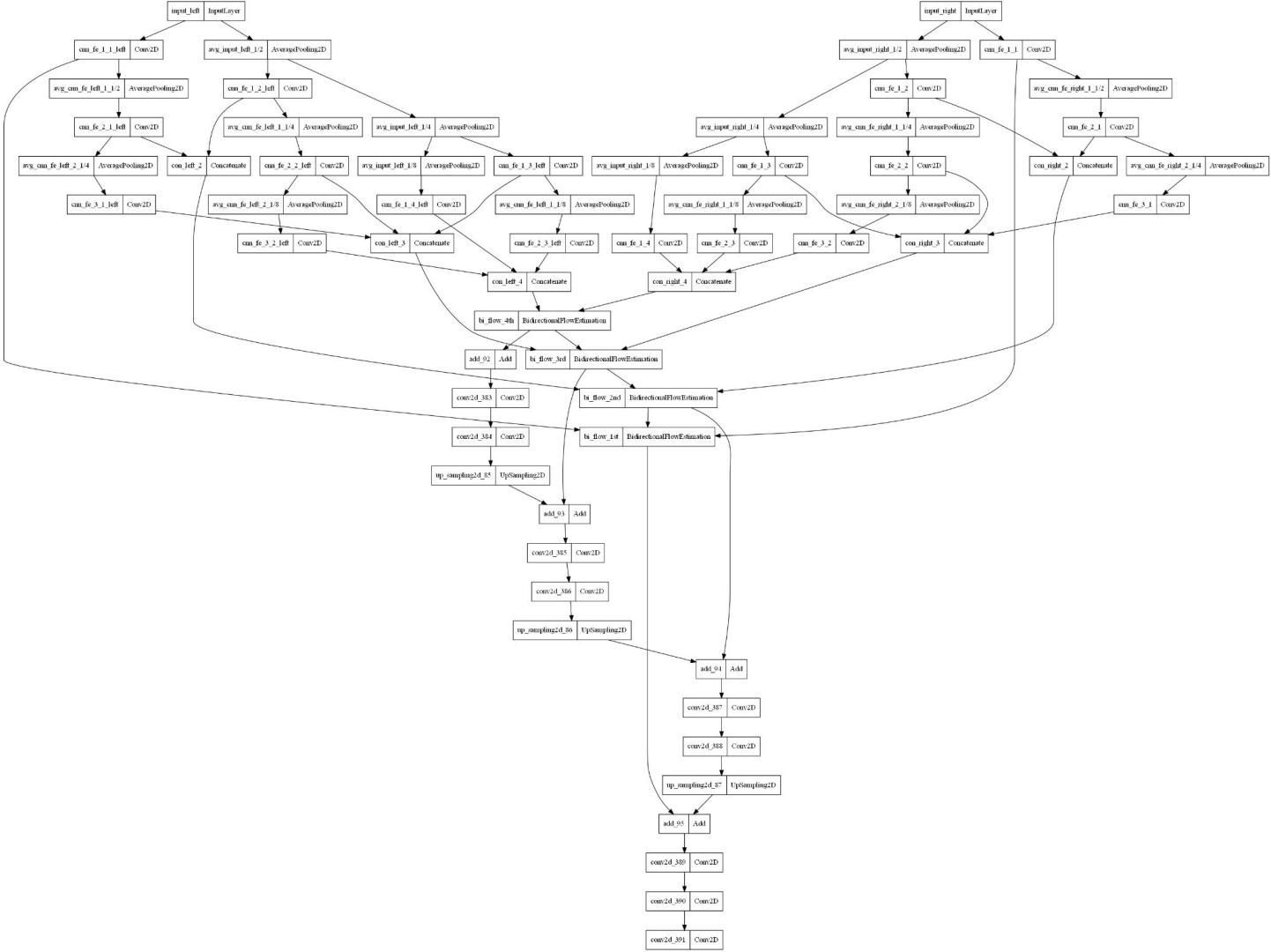




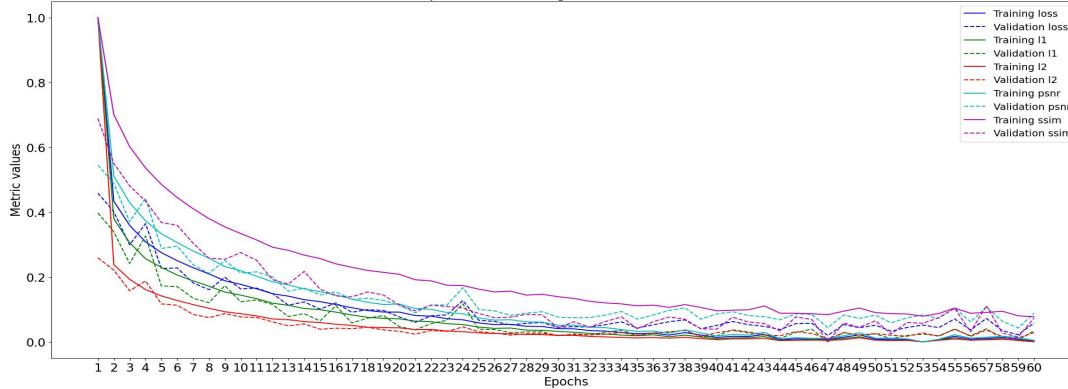
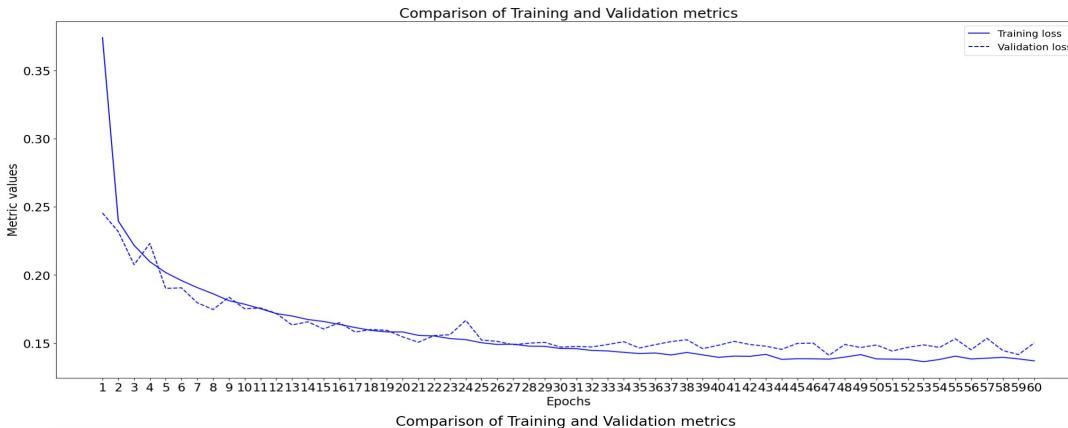
FBNet - decoder



FBNet

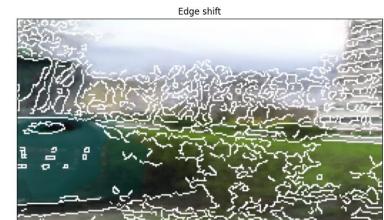
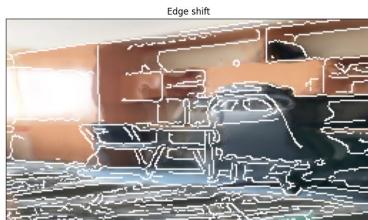


Initial model tests



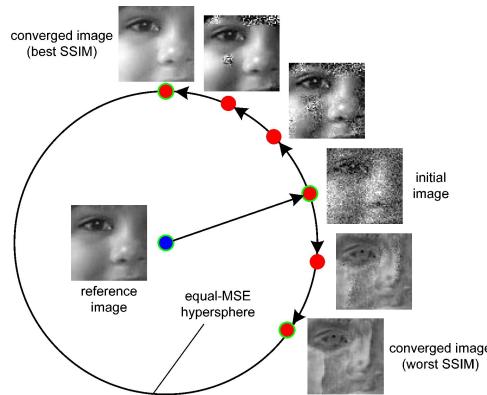


Initial model tests

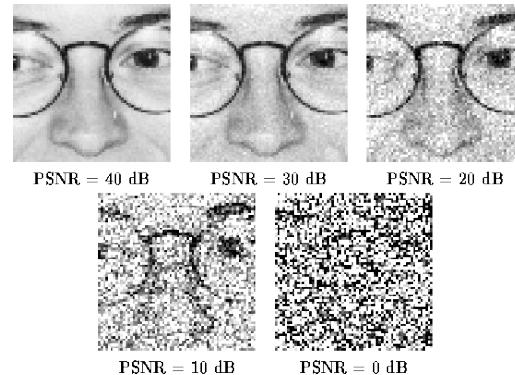


Loss function

SSIM:



PSNR:



$$\text{L1: } |\mathbf{x}|_1 = \sum_{r=1}^n |x_r|.$$

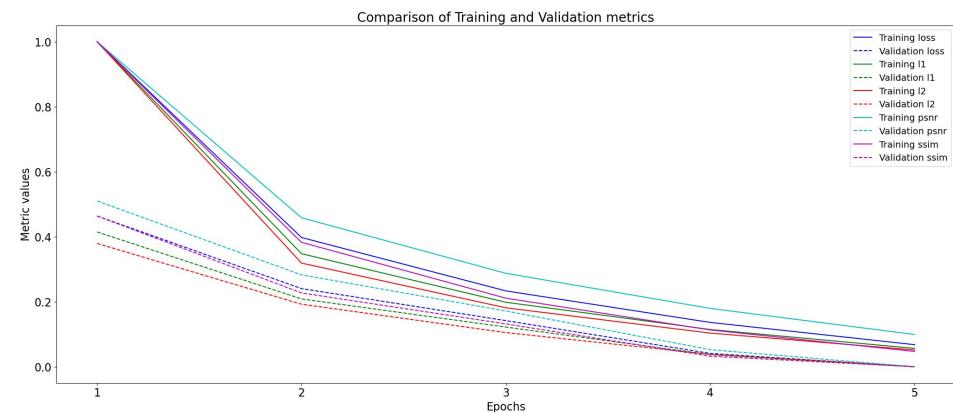
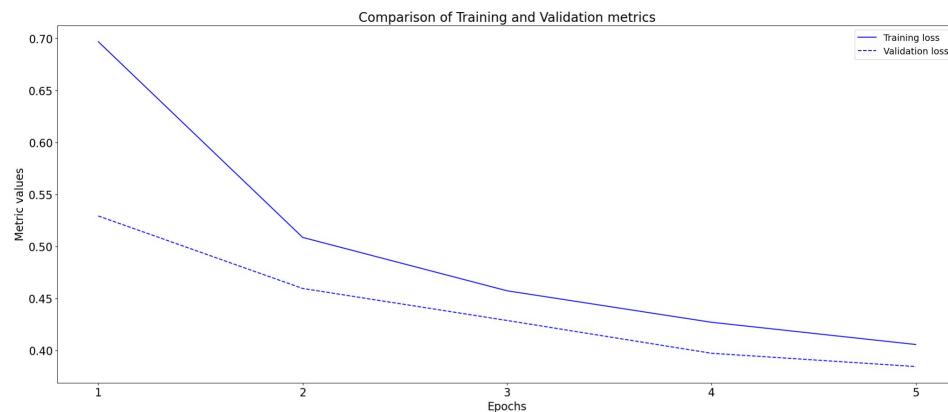
$$\text{L2: } |\mathbf{x}| = \sqrt{\sum_{k=1}^n |x_k|^2},$$

$$\text{Loss} = \text{SSIM} + \text{PSNR} + 5.0 * \text{L1} + 10.0 * \text{L2}$$



Trening

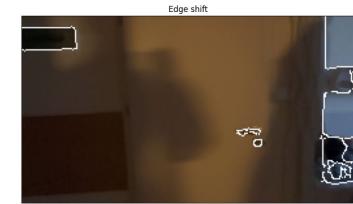
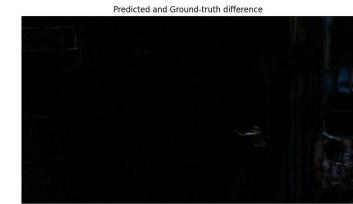
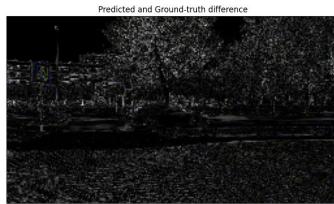
- TFRecords for processing 19000 training samples (25.5 GB)
- Nadam optimizer ($lr = 0.0001$, $clipvalue=1.0$, $clipnorm=1.0$)
- Epoch time: 22 minutes (1xGTX970 4GB, i5 4690K)
- Model parameters: 1 199 971



Results

Results for a test set of Vimeo90K in resolution 144 x 256 px

- PSNR: 30.44 (SOTA: 36.76)
- SSIM: 0.9096 (SOTA: 0.9800)



Results

First frame



Predicted and Ground-truth difference



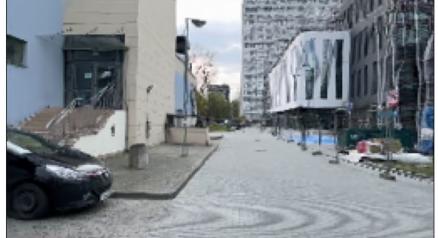
First frame



Predicted and Ground-truth difference



Predicted frame



Ground-truth frame



Predicted frame



Ground-truth frame



Second frame



Edge shift



Second frame



Edge shift



Results

First frame



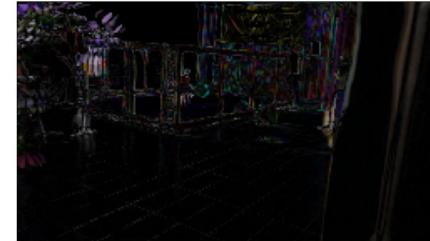
Predicted and Ground-truth difference



First frame



Predicted and Ground-truth difference



Predicted frame



Ground-truth frame



Predicted frame



Ground-truth frame



Second frame



Edge shift



Second frame



Edge shift



Model issues

First frame



Predicted and Ground-truth difference



First frame



Predicted and Ground-truth difference



Predicted frame



Ground-truth frame



Predicted frame



Ground-truth frame



Second frame



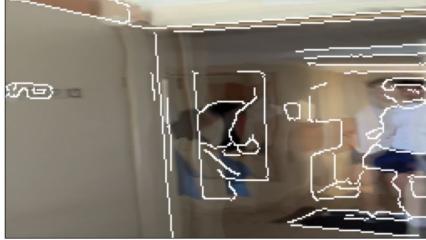
Edge shift



Second frame



Edge shift

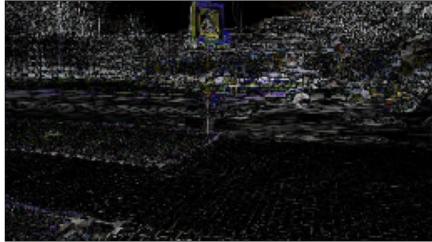


Model issues

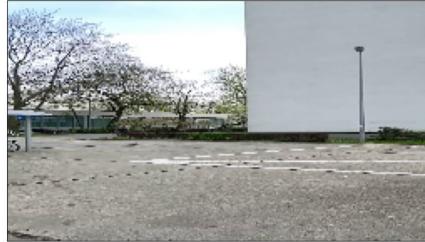
First frame



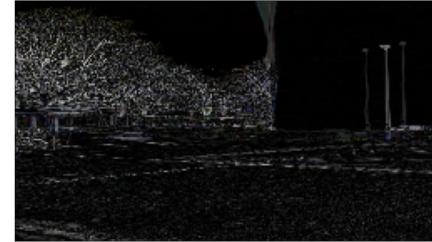
Predicted and Ground-truth difference



First frame



Predicted and Ground-truth difference



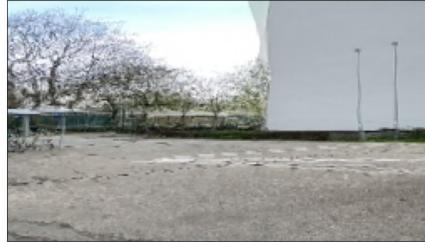
Predicted frame



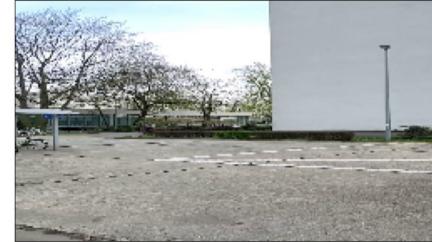
Ground-truth frame



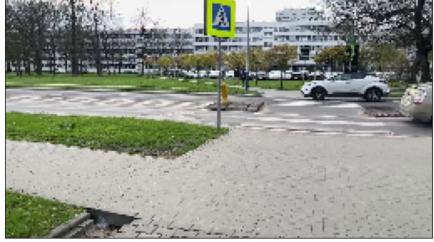
Predicted frame



Ground-truth frame



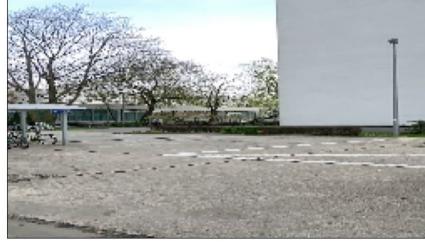
Second frame



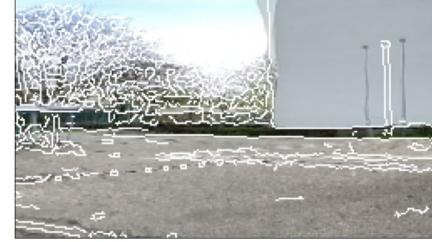
Edge shift



Second frame



Edge shift





Project development

- Teach the model on a dataset with a bigger motion
- Fix described model issues
- Create an environment for an end-to-end frame-boosting application



Resources

1. Single Image Super Resolution with deep convolutional neural networks
2. [Real-Time Intermediate Flow Estimation for Video Frame Interpolation](#)
3. [Depth-Aware Video Frame Interpolation](#)
4. [BiFormer: Learning Bilateral Motion Estimation via Bilateral Transformer for 4K Video Frame Interpolation](#)
5. [Attention is all you need](#)
6. [Video Frame Interpolation via Adaptive Convolution](#)
7. [Large Motion Frame Interpolation](#)
8. [FILM: Frame Interpolation for Large Motion](#)
9. [Multi-view Image Fusion](#)
10. [Perceptual Losses for Real-Time Style Transfer and Super-Resolution](#)
11. [Image Style Transfer Using Convolutional Neural Networks](#)
12. [Exploring Motion Ambiguity and Alignment for High-Quality Video Frame Interpolation](#)
13. [PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume](#)