

# Project Proposal:

## Video Frame Interpolation using Pyramid Representation

Jan Murić, Paulo Erak

This proposal introduces our project: “Video Frame Interpolation Using Pyramid Representation.” The primary goal of this project is to generate intermediate video frames between two given input frames. By achieving this, we can enhance video quality in several ways, such as increasing frame rates, creating slow-motion effects, improving video compression, and generating animations from fewer frames.

Video frame interpolation has significant applications in entertainment, media editing, and machine learning tasks, making it an essential research area in computer vision.

### Survey of Literature

To provide context and background, we will examine state-of-the-art approaches to video frame interpolation. Specifically, we will focus on:

- **Pyramidal Representations:** Widely used for multi-scale processing to enhance feature extraction and interpolation accuracy [2][4][5].
- **Convolutional Models:** Applied for extracting spatial and temporal features crucial for frame synthesis.
- **Optical Flow Estimation:** A commonly employed method for estimating pixel displacements between frames [1][3][4].
- **Loss Functions:** Used to optimize models for minimizing errors in synthesized frames.

The field of video frame interpolation has seen rapid advancements over the past eight years, with several influential papers proposing innovative architectures and methodologies. Datasets like Vimeo90k and Middlebury are frequently cited as benchmarks for evaluating interpolation models [1][2][3][4][6].

### Methodology

We will primarily use the Vimeo90k Triplet dataset, which consists of 3-frame sequences at a fixed resolution extracted from 15,000 video clips. This dataset is specifically designed for temporal frame interpolation tasks. In addition, we may include other publicly available video materials for qualitative demonstrations of our results.

Our proposed method involves the following steps:

1. **Pyramid Representation:** Construct a pyramidal representation of input frames.
2. **Feature Extraction:** Apply a Unet-like encoder as the feature extractor on each level of the pyramid.
3. **Optical Flow Estimation:** Estimate and refine optical flow at each pyramid level.

4. Feature Warping and Fusion: Warp features using the estimated flow, fuse the warped features, and decode them at each level.
5. Final Frame Refinement: Enhance the synthesized frame to minimize artefacts and improve visual quality.

While existing papers and implementations provide strong foundations, we aim to introduce innovations, particularly by trying to speed up the learning process as we want the video interpolator to be able to be trained on common consumer hardware.

## Evaluation Plan

We will evaluate the results of our project through both qualitative and quantitative measures:

- Qualitative Evaluation: We will perform visual inspection of the interpolated frame compared to the ground truth (original middle frame) to identify artifacts, inconsistencies, and blending quality.
- Quantitative Evaluation: We will use metrics such as Peak Signal-to-Noise Ratio (PSNR) to assess the accuracy of the synthesized frame.

By combining these methods, we will provide an analysis of our implementation's performance.

## References

- [1] Hahm, X. J. a. L. W. a. J. C. a. Y. C. a. J. K. a. C.-h., 2023. A Unified Pyramid Recurrent Network for Video Frame Interpolation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1578-1587.
- [2] Liu, S. N. a. F., 2020. Softmax Splatting for Video Frame Interpolation. *CoRR*, Volume abs/2003.05534, p. 10.
- [3] Meister, S. a. H. J. a. R. S., 2018. UnFlow: Unsupervised Learning of Optical Flow With a Bidirectional Census Loss. *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 32.
- [4] Reda, F. K. J. T. E. S. D. P. C. C. B., 2022. FILM: Frame Interpolation for Large Motion. *Computer Vision -- ECCV 2022*, pp. 250--266.
- [5] W. Shen, W. B. G. Z. L. C. X. M. a. Z. G., 2020. Blurry Video Frame Interpolation. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 10.
- [6] Wu, C. Z. a. J. L. a. J. T. a. G., 2023. Video Frame Interpolation with Densely Queried Bilateral Correlation. *ArXiv*, Volume @article{Zhou2023VideoFI,.