cuFLAVR

Aditya Hota, Richard Chen, Kaan Erdogmus
CIS 565 Fall 2021
Milestone 3 Presentation

FLAVR: Flow-Agnostic Video Representations for Fast Frame Interpolation

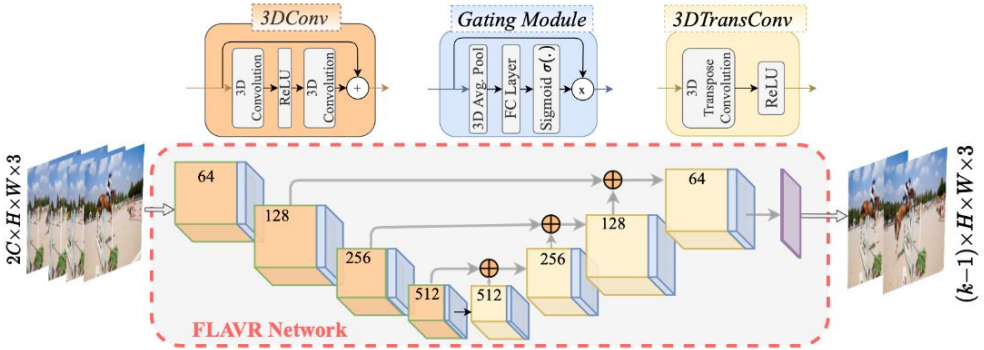
Tarun Kalluri * UCSD

Deepak Pathak CMU

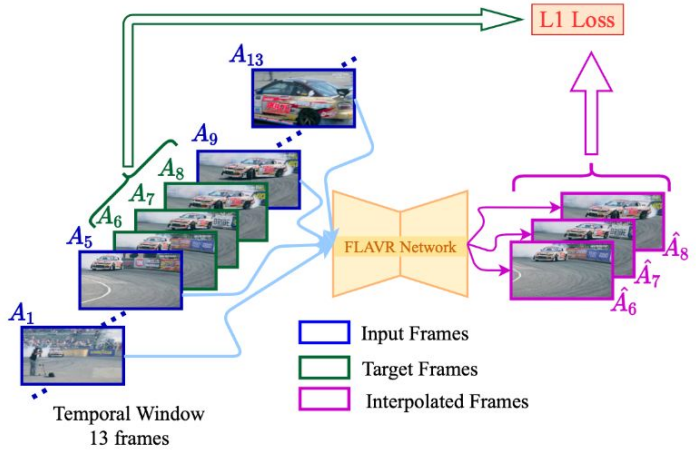
Manmohan Chandraker UCSD

Du Tran Facebook AI

https://tarun005.github.io/FLAVR/



(a) Overview of the proposed architecture



(b) Sampling procedure

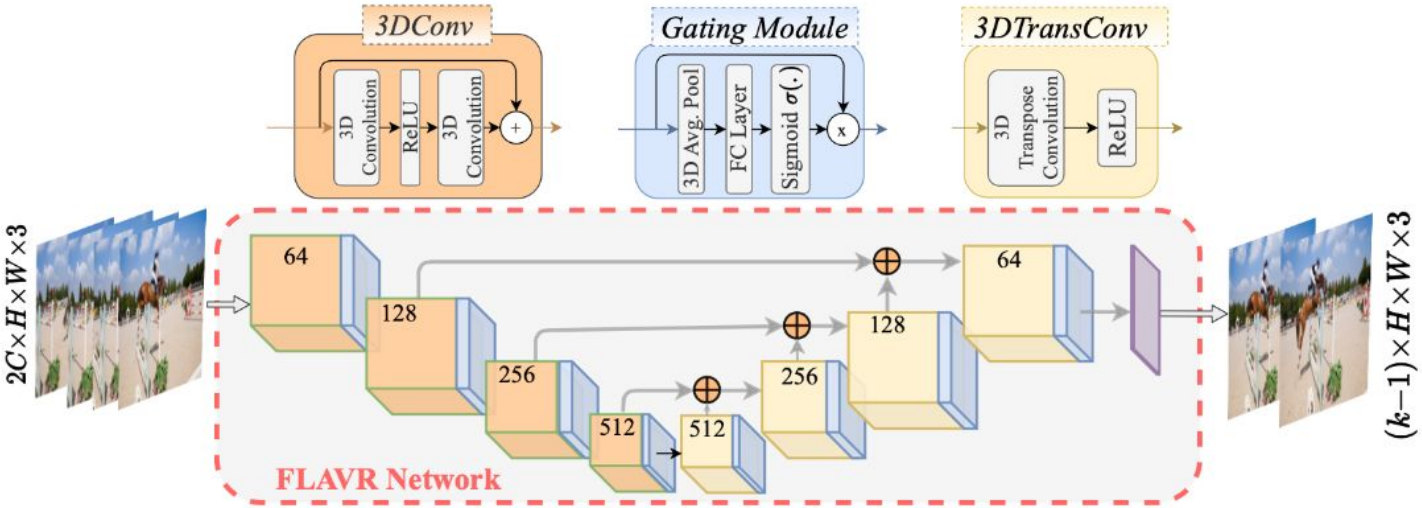
Milestones

- Milestone 3 (12/06)
  - Finish layers for 3D convolutions and gating
  - Match PyTorch model weights with our cuFLAVR network
  - Convert video into image frames (inference inputs)
- Final Deliverable (12/12)
  - Combine all custom layers together
  - Automated pathway for generating interpolated videos
  - Performance analysis (and potential comparison to PyTorch)

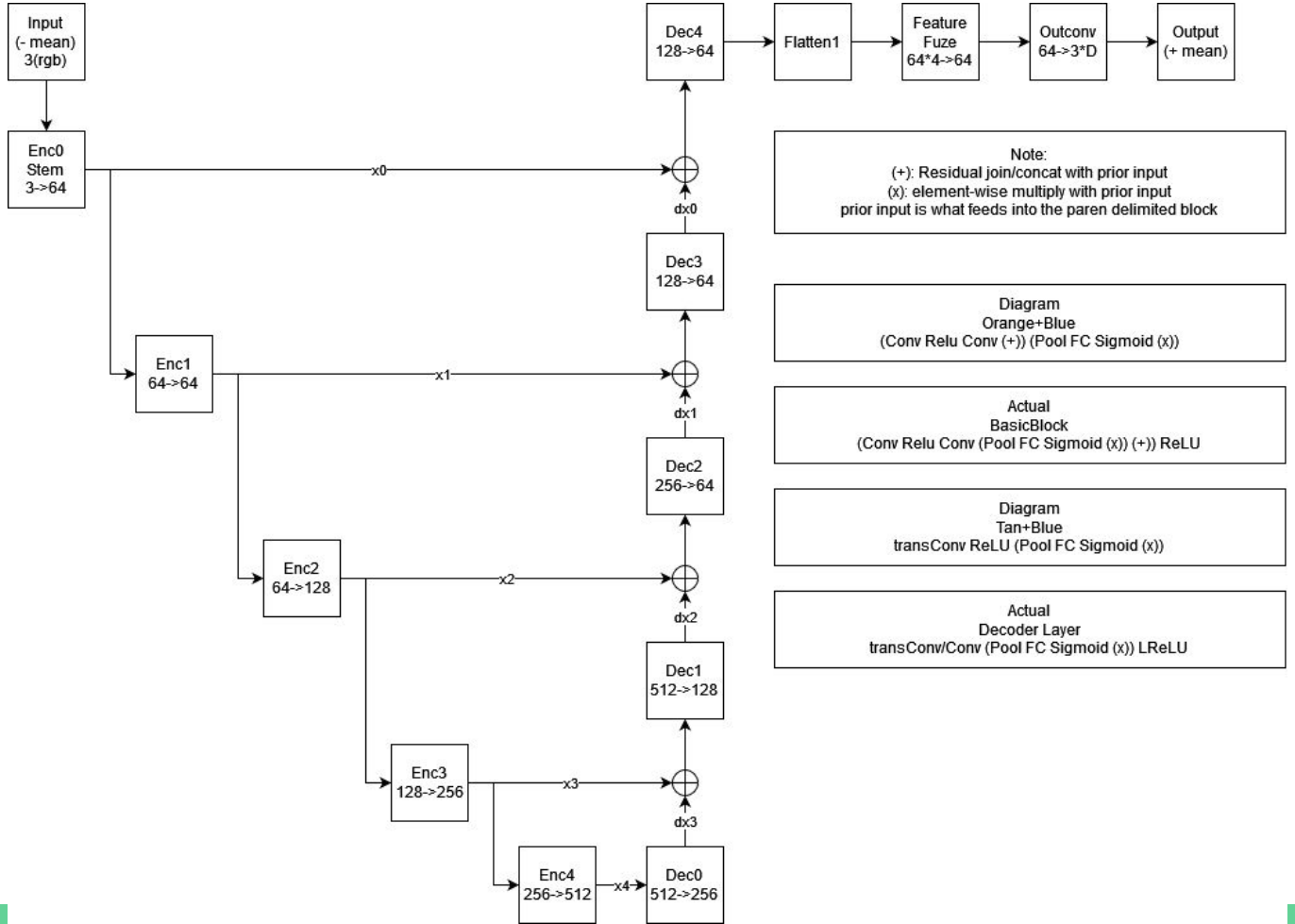
Completed Work

- Milestone 3 (12/06)
  - ⚠️ Reviewed network details and noticed discrepancies with paper
    - Diagrams and implementation details in paper are not accurate
  - ✅ Understood which convolutions are needed for cuDNN implementation
  - ✅ Implemented 3D convolutions
  - ✅ Implemented ReLU and Sigmoid cuDNN layers
- Other progress
  - ✅ Created wrapper for NN layers to simplify cuDNN calls
  - Combining layers is in progress

Model Revision (old)

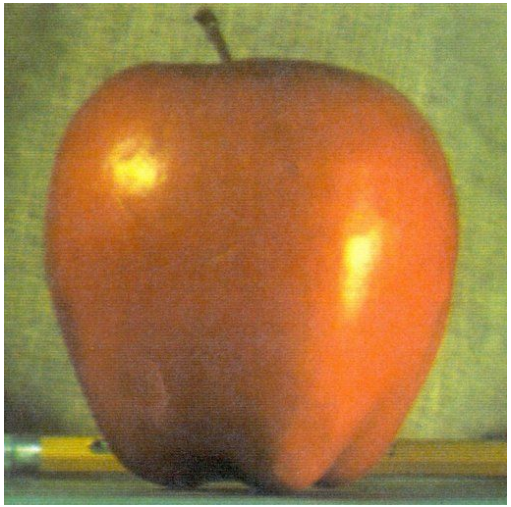


Model Revision (correct)



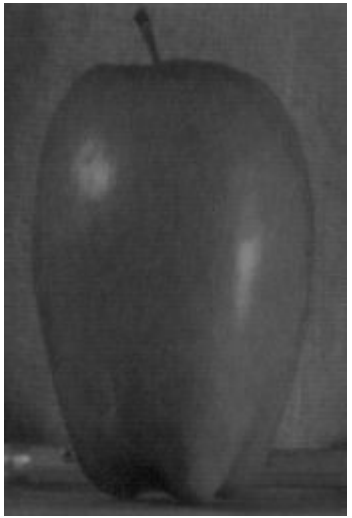
3D Convolutions using cuDNN

- Implemented all code required to setup and run convolution
- Values match expected outputs from PyTorch!
  - Checked float values for matching; dark color due to PIL in Colab



(a) Original image

Kernel: (1, 2, 2)
Stride: (1, 2, 3)



(b) PyTorch result

(c) cuDNN result

Model Simplifications

- Streamlined PyTorch model to remove unused code
- Exported info for all layers of model in plain text
- Able to do all convolutions without batch normalizations

Next Milestone

- Final Presentation (12/13)
  - Finish as many remaining CNN layers as possible
    - Integrate into PyTorch if needed
  - Load all weights from PyTorch model
  - Load frames into CUDA application
  - Generate output
  - Performance analysis
  - Update the README

References

Papers:

Kalluri, T., Pathak, D., Chandraker, M., & Tran, D. (2020). Flavr: Flow-agnostic video representations for fast frame interpolation. *arXiv preprint arXiv:2012.08512*.

Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., & Paluri, M. (2018). A closer look at spatiotemporal convolutions for action recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 6450-6459).

2D Convolutions Guide:

Peter Goldsborough: 2D Convolutions using cuDNN

Q&A