Deep Learning For Videos

Anil Chandra Naidu Matcha

TensorFlow User Group Mysuru

GDG Bangalore

Co-Learning Lounge

APPLIED SINGULARITY

Presents

DEEP LEARNING FOR VIDEOS

Anil Chandra
Naidu Matcha
Co-Founder at Vadoo backed by
Entrepreneur First

On: 30th May 2020
Time:4:00pm-5:00pm (IST)

BIO

Co-Founder at Vadoo

Coach at Upgrad, Board Infinity

Ex-Samsung, Cisco

Contact :- https://www.linkedin.com/in/anilmatcha/

AGENDA

What & Why

Applications

Architectures

Datasets

Q&A

What & Why

Deep learning revolution started in 2012 💥💥

Alexnet beat all the existing algorithms in ILSVRC 2012 👊🏽👊🏽

Now deep learning is taken for granted on all image related tasks 💪💪

A new trend is emerging recently🚀 🚀

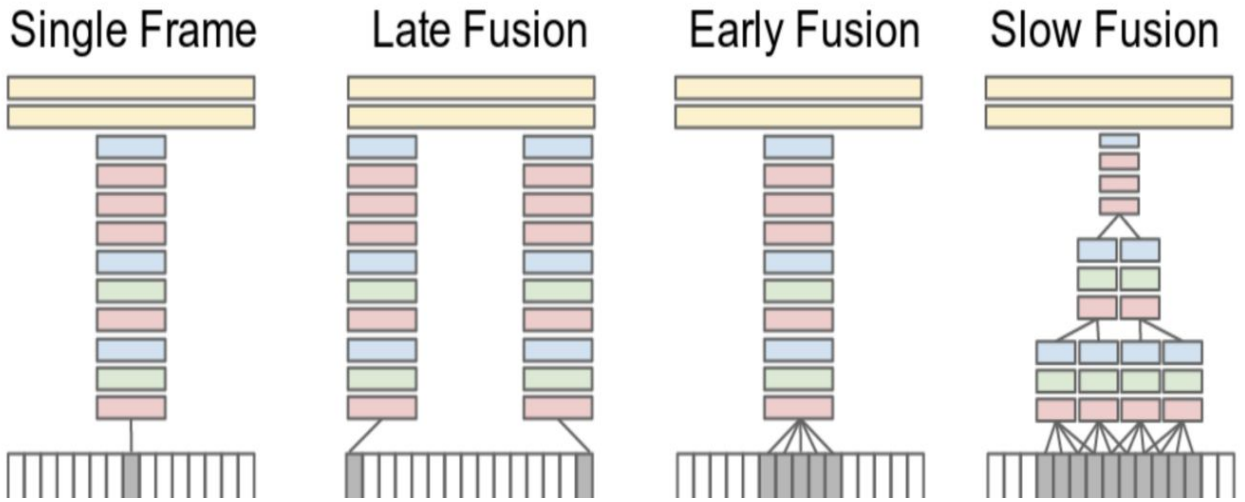
We are not limited to images anymore 📷📷

Now machines can understand Videos 🎥🎥

Action Recognition

Single Stream Neural Networks

Failed to capture temporal information



Problems

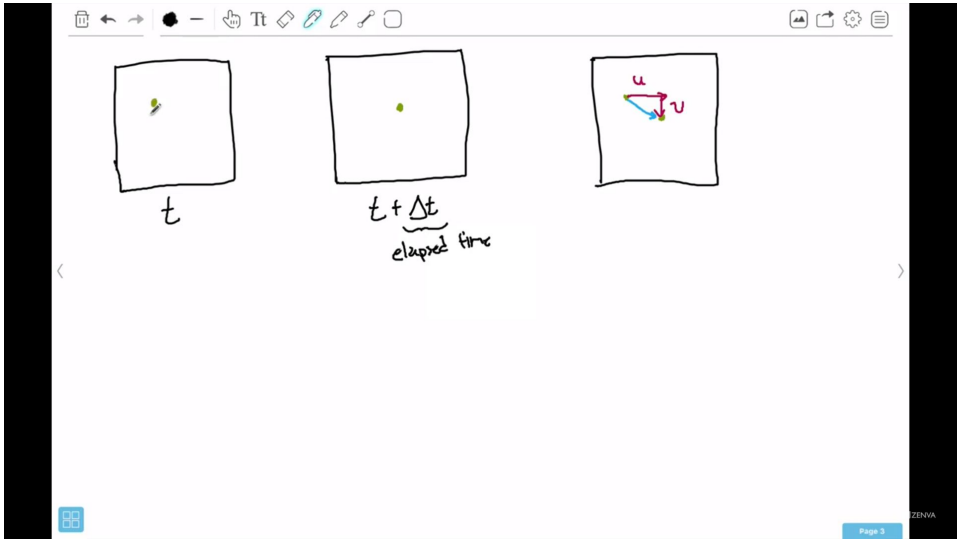
Didn’t provide any improvement compared to using just a single frame

Learnt spatio-temporal features didn’t capture motion information

Dataset used was not diverse and hence couldn’t learn the information

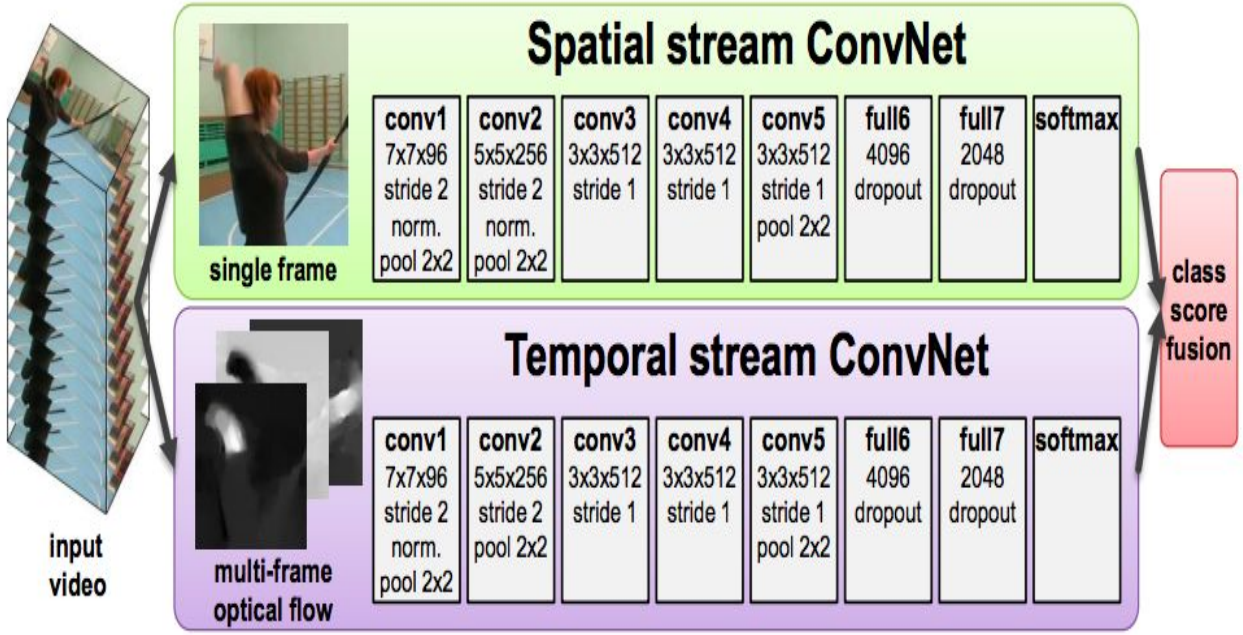
Optical Flow

Track the flow of pixels across consecutive frames



Two Stream Neural Networks

Average of outputs



Problems

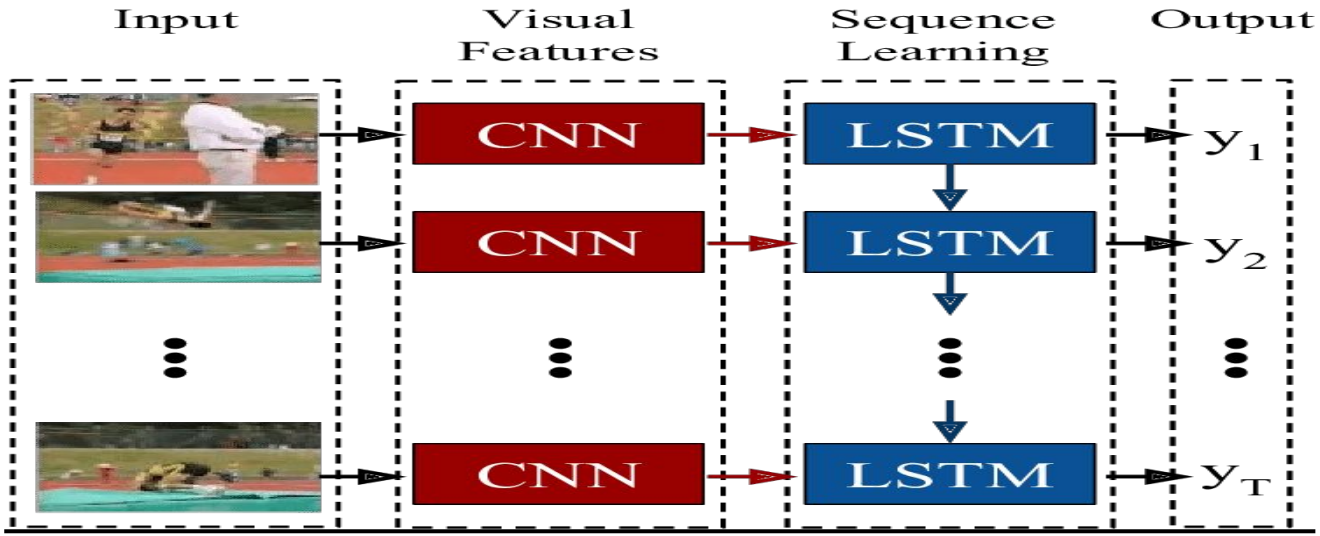
Need to pre-compute optical flow

Not end-to-end trainable

No transfer learning on temporal stream convnet

LRCN

End-to-end trainable

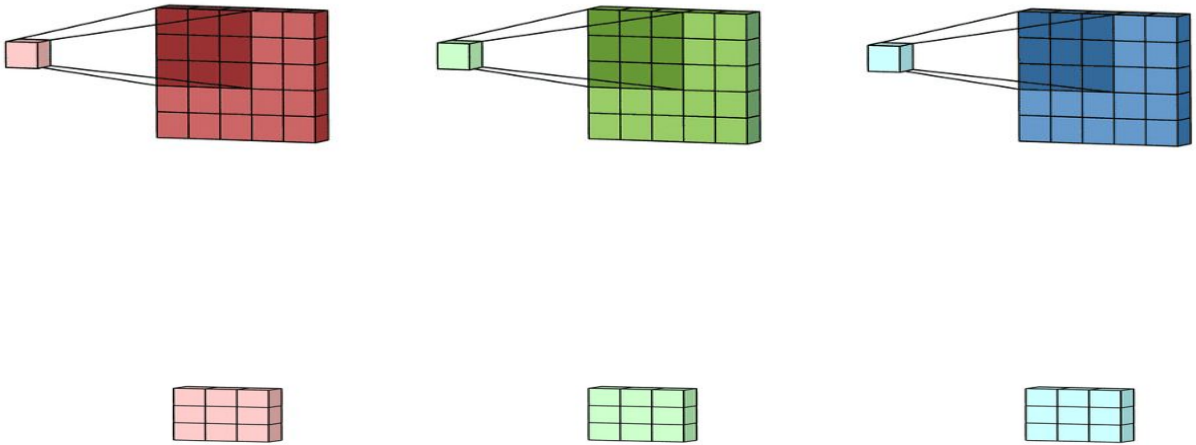


Problems

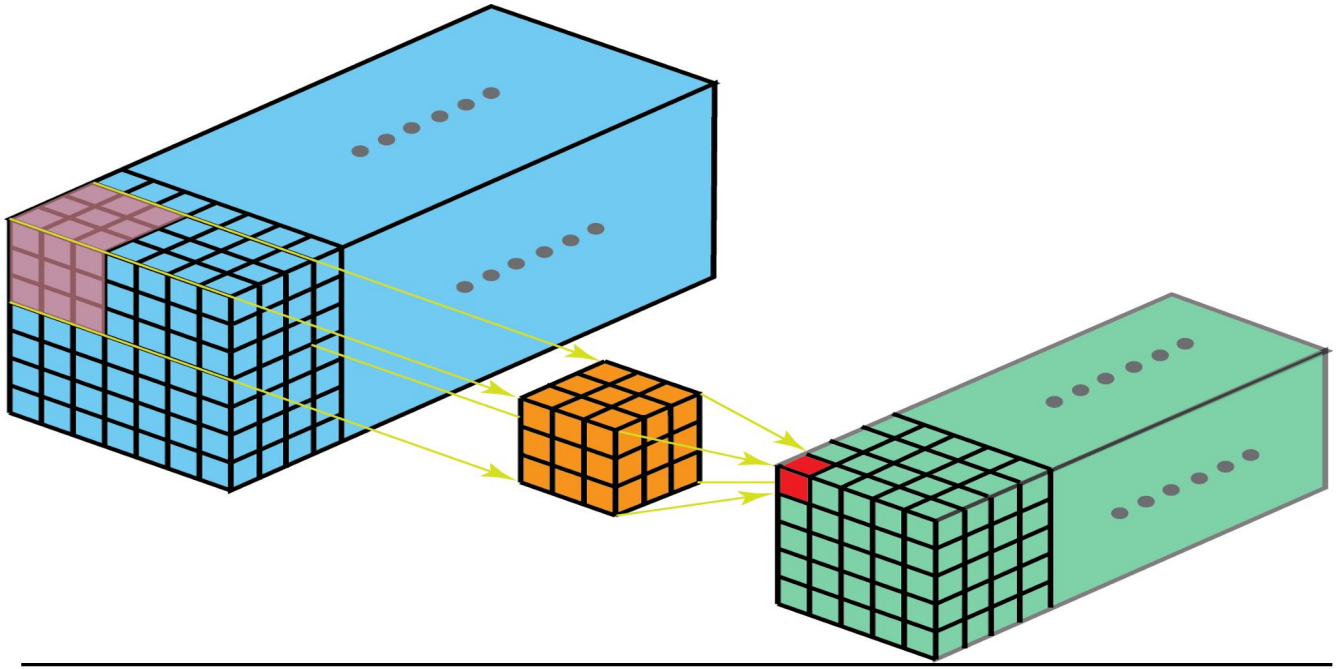
Inability to capture long-term temporal information

Needed pre-computing if optical flow is used as input

2D convolution

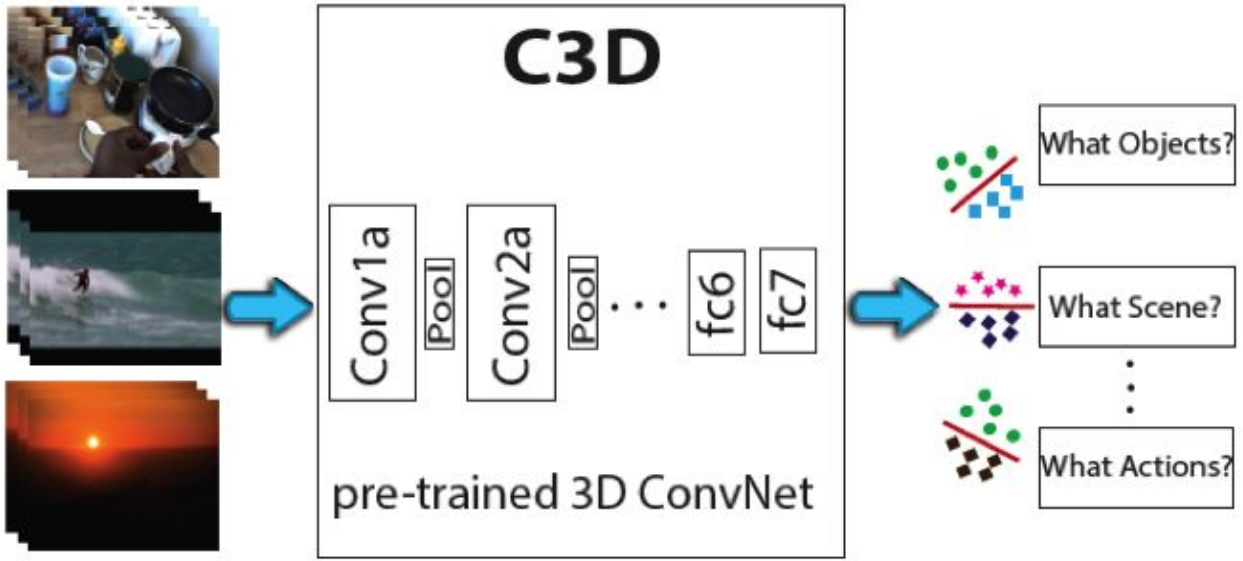


3D convolution



C3D

3x3x3 Convolution
2x2x2 Pooling

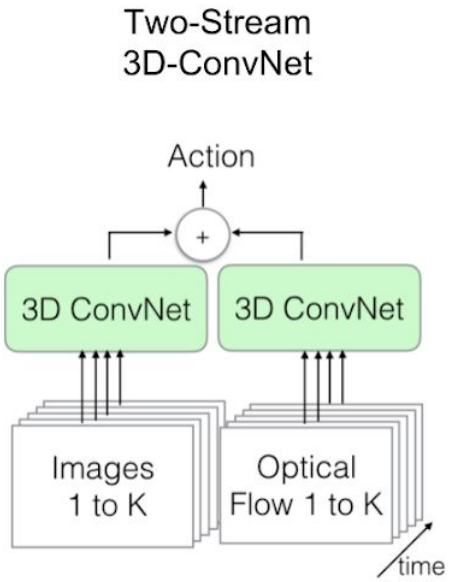


I3D

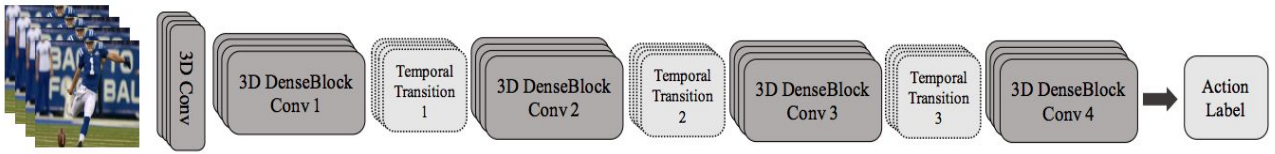
3D ConvNets for two stream architecture

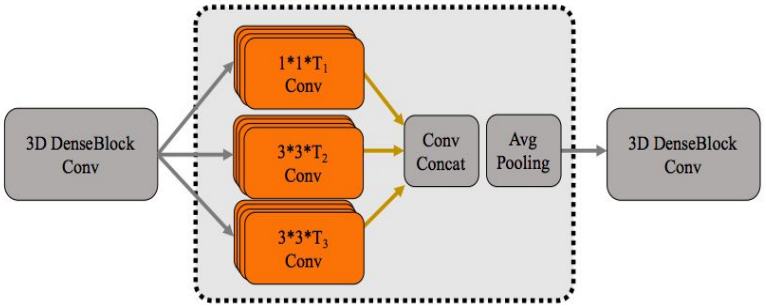
Pre-trained models from 2D extended to 3D

Multiple frames passed to spatial stream now



T3D





3D Temporal Transition Layer

T3D

DenseNet architecture with 3D convolution

Temporal Transition layer to model variable temporal depths similar to Inception network for 2d

Problems

More computational power required

More number of parameters due to additional 3rd layer in filter

Datasets

HMDB51

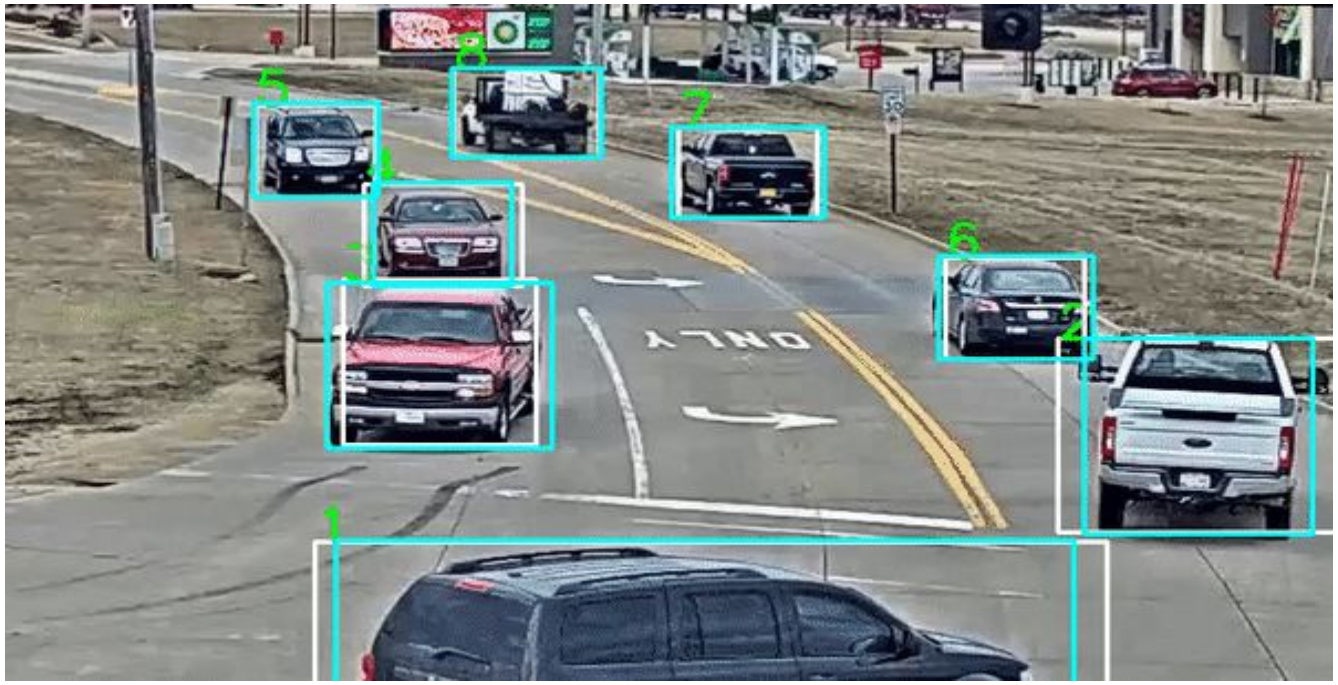
UCF-101

Sports-1M

Youtube-8M

Kinetics Dataset

OBJECT TRACKING



DIFFICULTIES

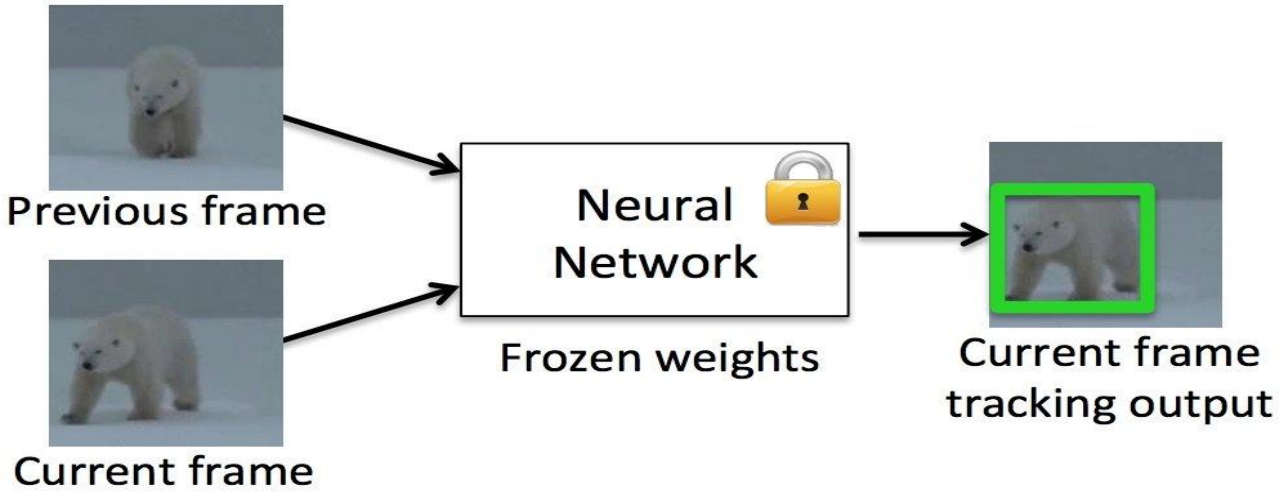
Occlusion

Identity switch

Moving camera

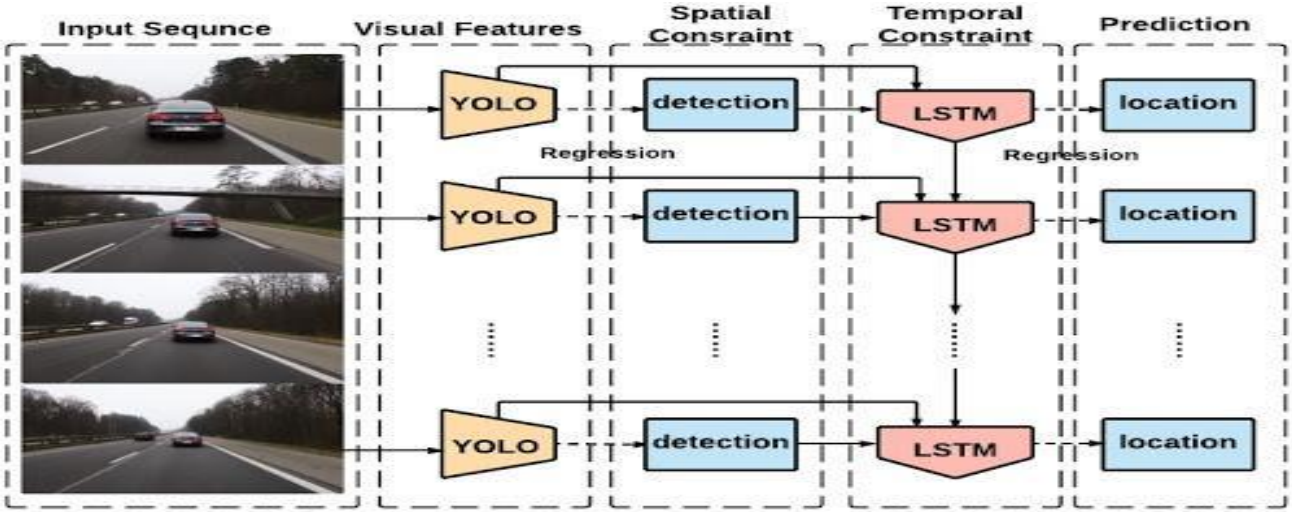
GOTURN

Trained on a lot of objects. Can track unseen objects



ROLO (Recurrent YOLO)

YOLO features and bounding box outputs combined to provide as input for LSTM



DEEPSORT

Use a network trained on person re-identification dataset to give a 128 vector for every object detection

Use this to compute distance between frames to track objects

Datasets

Stanford Drone Dataset

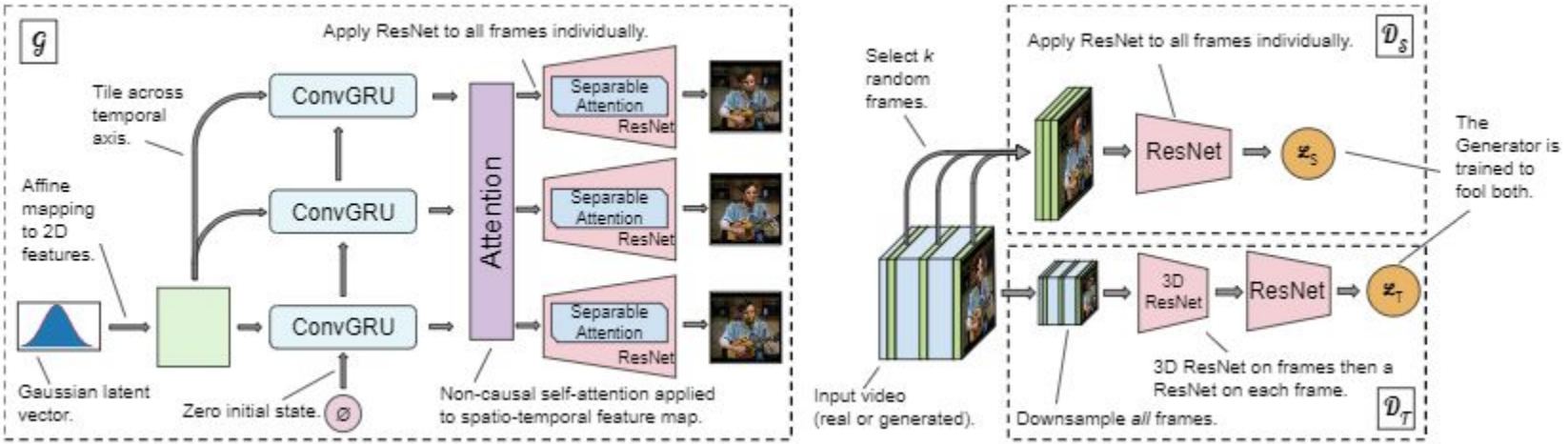
MOT17

PETS 2017

GENERATING VIDEOS

GAN for videos

DVD-GAN



Super-resolution for videos

Super resolution is the task of upscaling low res inputs to high res using a neural network

Direct applying of super-res on every frame in a video results in jarring artifacts

FRVSR

1. Has two networks Fnet and SRnet

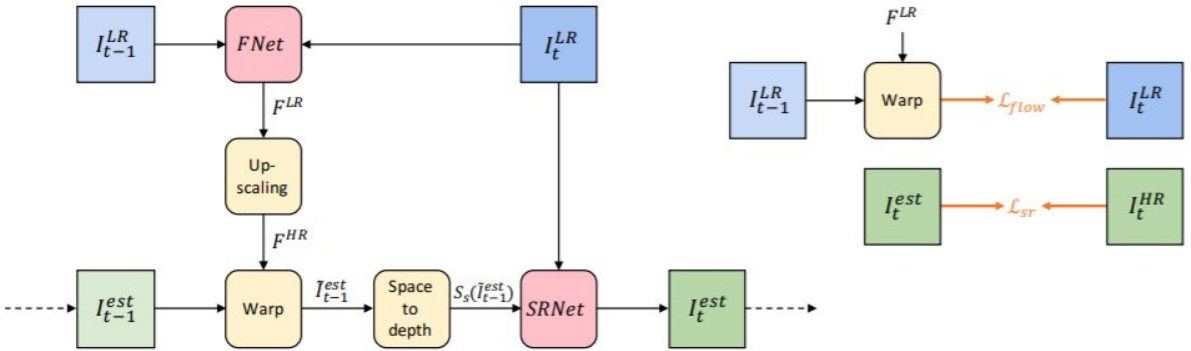


Figure 2: Overview of the proposed FRVSR framework (left) and the loss functions used for training (right). After computing the flow F^{LR} in LR space using FNet, we upsample it to F^{HR}. We then use F^{HR} to warp the HR-estimate of the previous frame I_{t-1}^{est} onto the current frame. Finally, we map the warped previous output $\tilde{I}_{t-1}^{\mathrm{est}}$ to LR-space using the space-to-depth transformation and feed it to the super-resolution network SRNet along with the current input frame I_t^{LR}. For training the networks (shown in red), we apply a loss on I_t^{est} as well as an additional loss on the warped previous LR frame to aid FNet.

Summary

We saw few applications of how deep learning can be applied for videos

Similarly there are many other works which are being done in video compression, video editing, video blending, vfx etc. and the future scope is limitless

Resources :-

C3D code

https://github.com/TianzhongSong/C3D-keras

UCF-101 Dataset

https://www.crcv.ucf.edu/data/UCF101.php