Super Resolution

Team Member Names:- VSR with ESTRNN: Lingyi Dong, Subia Ansari, Yuchen Tang;
VFI: Lin , Caleb; VSR with CNN: Laula, Scott; Blind SR: Deva

PURDUE UNIVERSITY | The Data Mine

Viasat

VIDEO SUPER RESOLUTION WITH CONVOLUTION NEURAL NETWORK(CNN)

Introduction

Propose a CNN that is trained on both the spatial and the temporal dimensions of videos to enhance their spatial resolution. Consecutive frames are motion compensated and used as input to a CNN that provides super-resolved video frames as output.

Method

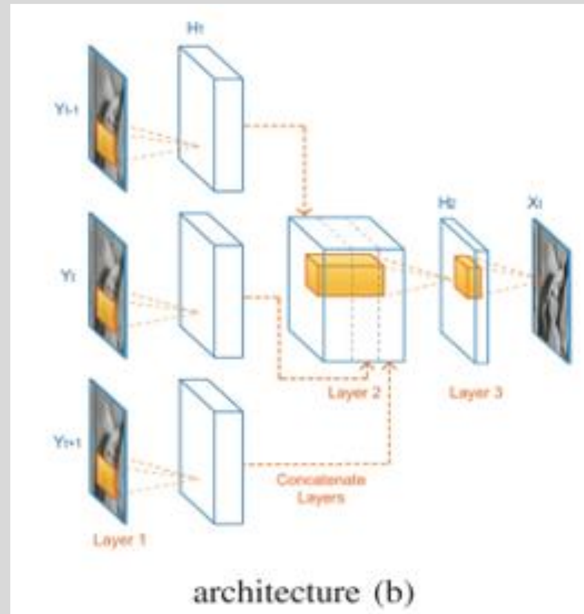
Weight Transfer from pretraining:

- Ensure Reference and video SR networks conditions to be equal
- Pretrained filter values from the reference model =>video SR models

Motion Compensation:

- Predict an image frame in a video, given the previous and future image frames by accounting for motion of the object in the video.

Architecture



architecture (b)

- Three convolutional layers, two hidden layers, H1 and H2 followed by a Rectified Linear Unit.
- Combine the frames after the first layer
- The output data of layer 1 is again concatenated along the first dimension and then used as input to layer 2.

Datasets

Train/Val : CVDL

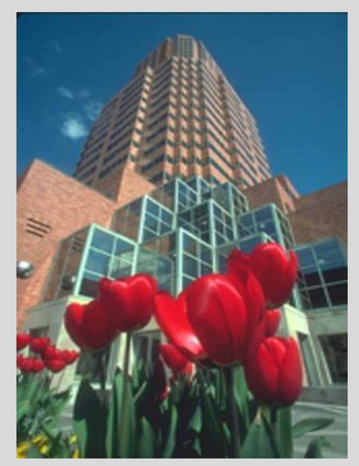
Test : | **Set 5** | **Set 14** | **BSD 100** | **Sun-Hays 80** | **Urban 100**

We can observe significant difference on some details such as the scarf on the women and the outline of the building while comparing Bicubic with SRCNN interpolation

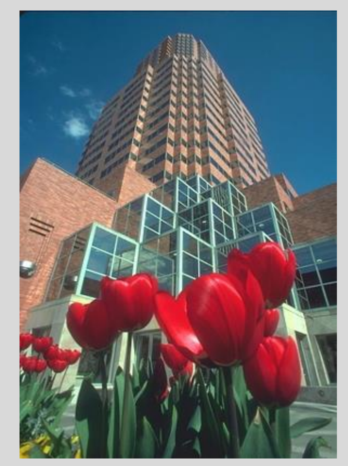
Results

Bicubic

SRCNN



Bicubic



SRCNN

VIDEO FRAME INTERPOLATION (VFI)

What is VFI?

Video frame interpolation (VFI) converts low frame rate (LFR) contents to high frame rate (HFR) videos by synthesizing one or more intermediate frames between given two consecutive frames

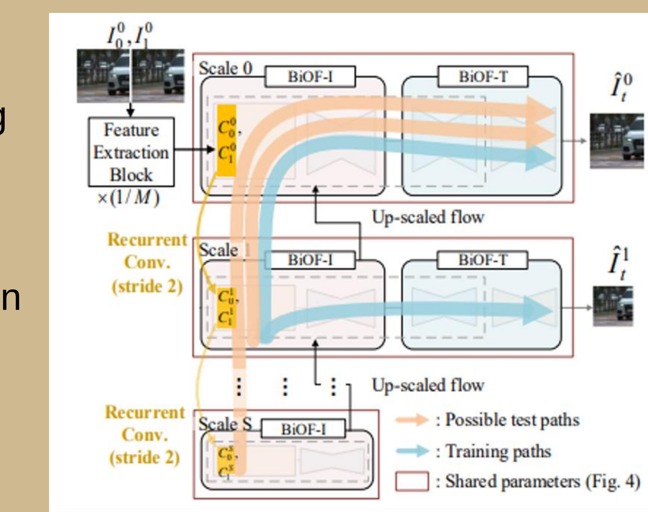
Applications

- Slow motion generation
- Frame rate upconversion
- Frame recovery in video streaming

The videos with a high frame rate can avoid common artifacts, and therefore are visually more appealing to the viewers.

Model - XVFI

- Propose a complementary flow reversal (CFR) that can generate stable optical flow estimation from time t to the input frames, boosting both qualitative and quantitative performances
- Can start from any downscaled input upward, which is adjustable in terms of the number of scales for inference according to the input resolutions or the motion magnitudes.



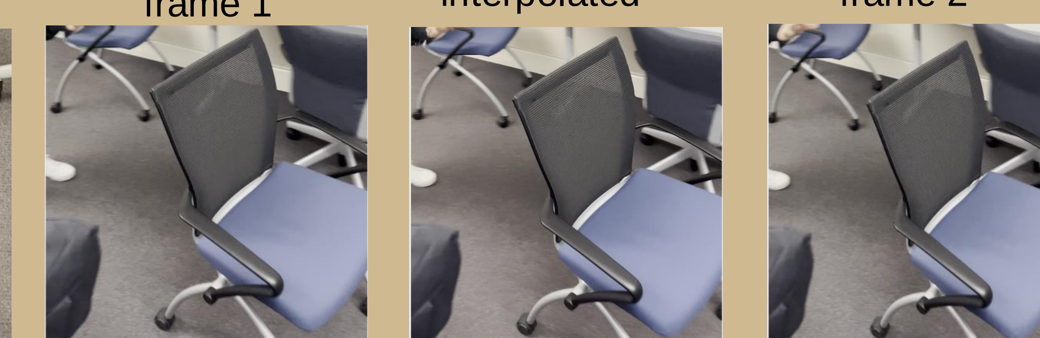
Results

Where it fails – fast moving object

frame 1 | interpolated | frame 2

Where it works – rigid object with small relative movement

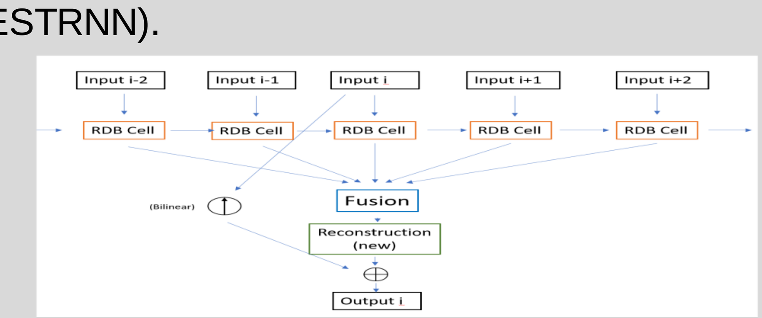
frame 1 | interpolated | frame 2



VIDEO SUPER RESOLUTION (VSR) WITH ESTRNN

Introduction and Background

ESTRNN is an efficient model used for deblurring videos, which is similar to VSR task. Information from neighboring image frames, also known as the Spatio-temporal relation, is critical VSR. We base our model on Efficient Spatio-Temporal Recurrent Neural Network (ESTRNN).



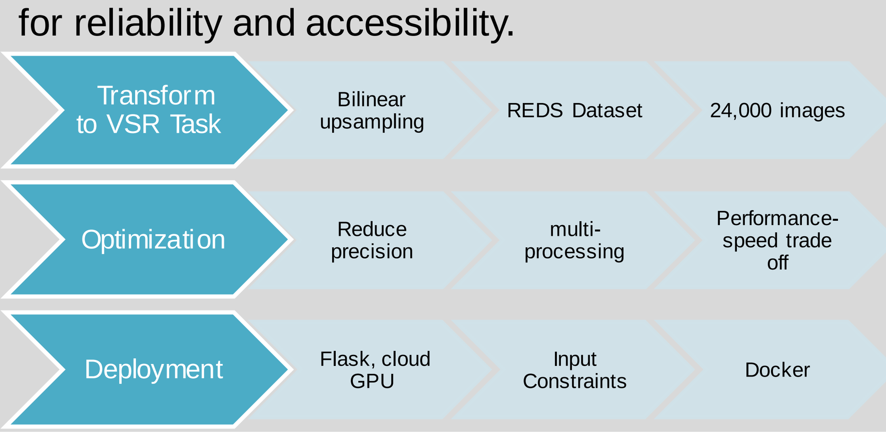
Research Method

Model Structure and Training

We proposed an upsampling re-constructor and a global skip connection with bilinear upsampling beyond the original ESTRNN structure for VSR task (as shown in the figure above). The new model was trained with REDS dataset (240 videos, ~24,000 images)

Model Optimization and Deployment

- **Optimization:** Implemented mixed precision method and boosted the frames per second by 50%.
- **Deployment:** Containerize the application using docker for reliability and accessibility.



Results

- **Top images:** original inputs upsampled through traditional methods (bicubic interpolation) by a factor of 4.
- **Bottom images:** super resolved outputs from the model by a factor of 4.

Future Work: Model Deployment

The key step to gain operational value from this model will be deploying the model in a live environment as a web-based application. In this project, we use Flask to wrap a mini-VSR model built with PyTorch, in the interest of providing the ability to perform VSR on short clips free of charge. The web-application is currently under construction and captures live video stream using built-in camera and converts the stream to jpeg images. We will also apply brightness filters to the video.

References & Acknowledgements

We thank our mentors Mark and Rafael from ViaSat for the guidance along the way and the DM for providing GPU assess on Purdue Gilbreth Computing Cluster.

This work is built on ESTRNN: https://github.com/zzh-tech/ESTRNN

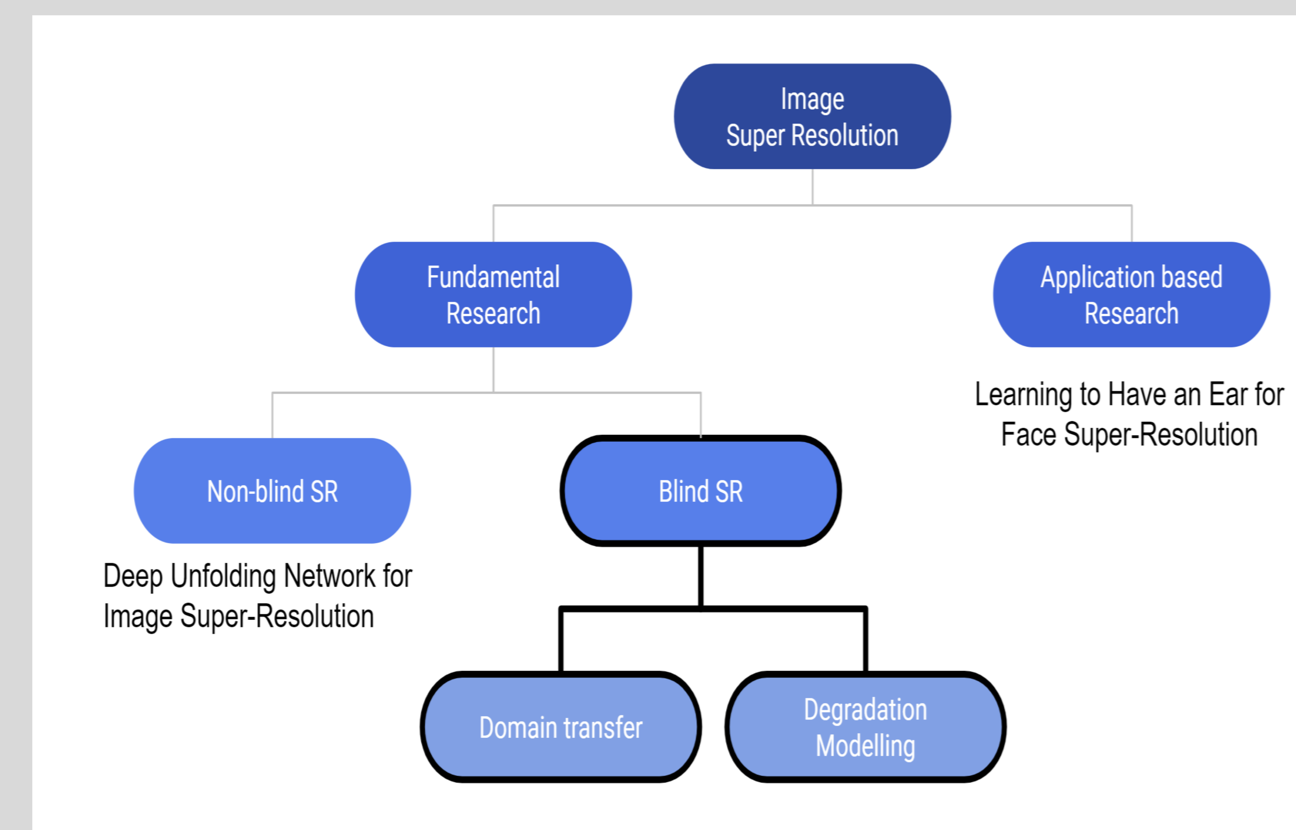
BLIND SUPER RESOLUTION

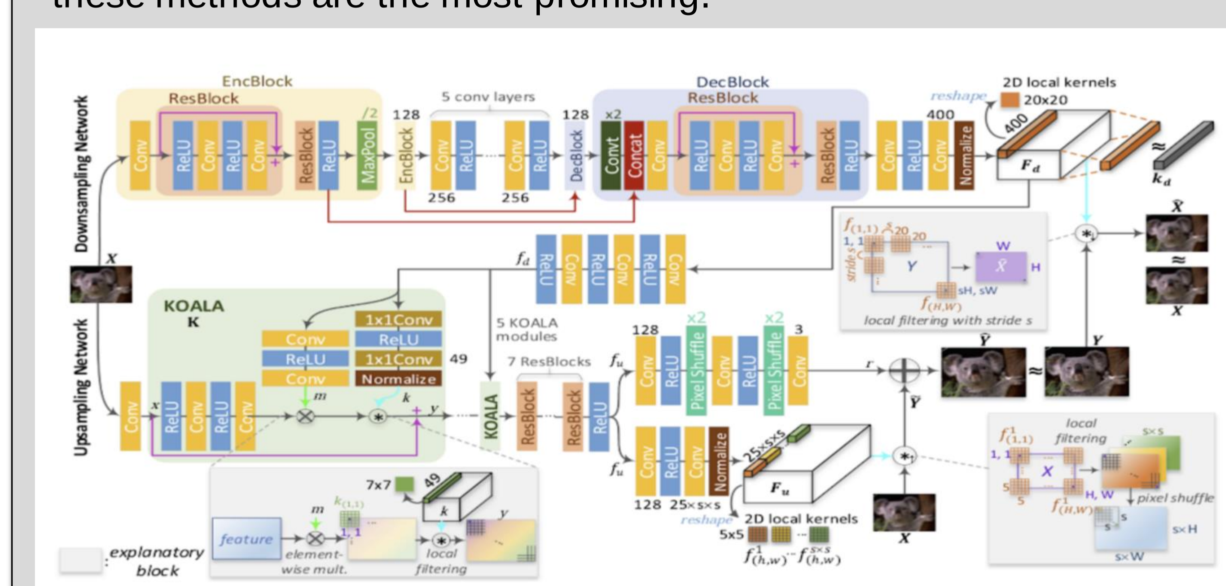
Introduction

The blind super-resolution aims to solve the problem of super resolving images which unknown degradations. The state-of-the-art super-resolution networks are trained using Bicubically downsampled I.e images that are downsampled using bicubic interpolation. This creates model bias and the networks get overfit towards this type of super-resolution. And when we try to deploy these trained networks in a real-life scenario, they fail because of this false assumption that all low-resolution images are downsampled using bicubic interpolation.

Method

There are several ways to solve this problem depending on different scenarios. The overall research is summarized in the taxonomy diagram shown on the right side, The most promising method out of these is degradation modeling-based methods. These methods can estimate degradation from any image and that is why these methods are the most promising.





Model

We present the state-of-the-art model for Blind super-resolution called KOALANet. This model overcomes several disadvantages over the previous methods. It can handle complex degradations as well as to adapt to new and unseen degradations very well because of the dynamic kernels. The idea is to generate the actual degradation kernels instead of just the degradation representation unlike the previous methods. These dynamic kernels are used in both, degradation estimation network and up sampling network.

Reference: https://github.com/hjSim/KOALANet