Review Series of Recent Deep Learning Papers:

Parameter Prediction Paper: Dynamic Filter Networks

Bert De Brabandere, Xu Jia, Tinne Tuytelaars, Luc Van Gool

NIPS 2016

Reviewed by : Arshdeep Sekhon

[1]Department of Computer Science, University of Virginia
https://qdata.github.io/deep2Read/

August 25, 2018

Motivation

1. Given related views humans can predict the next view
2. Given a video frame humans can predict the next frame

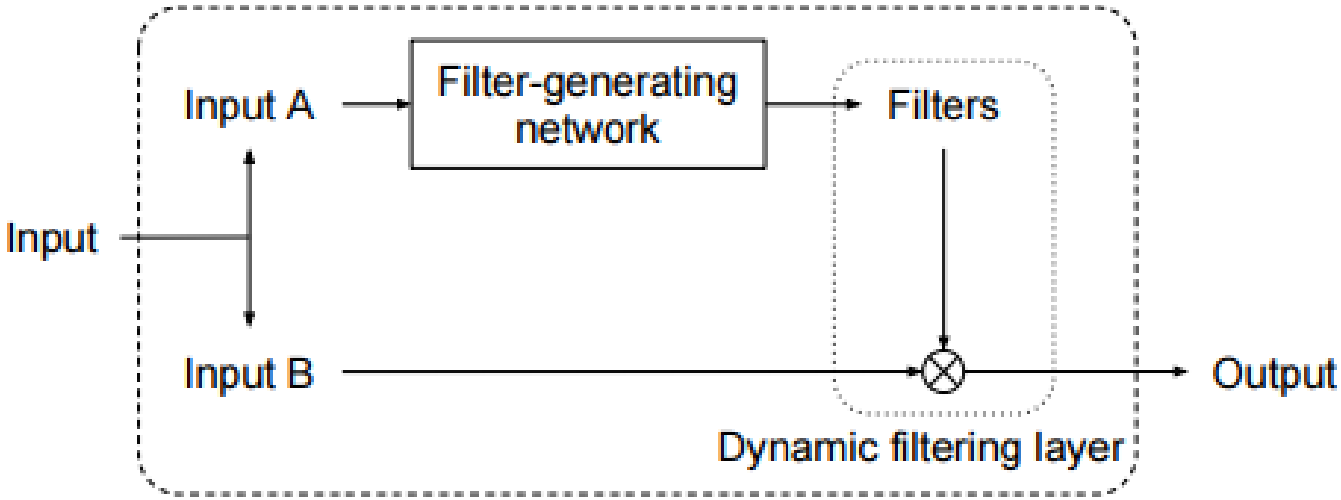
Motivation

1. Given related views humans can predict the next view
2. Given a video frame humans can predict the next frame
3. Deep Networks trained for the same tasks use the same trained filtering operation for all samples.

Motivation

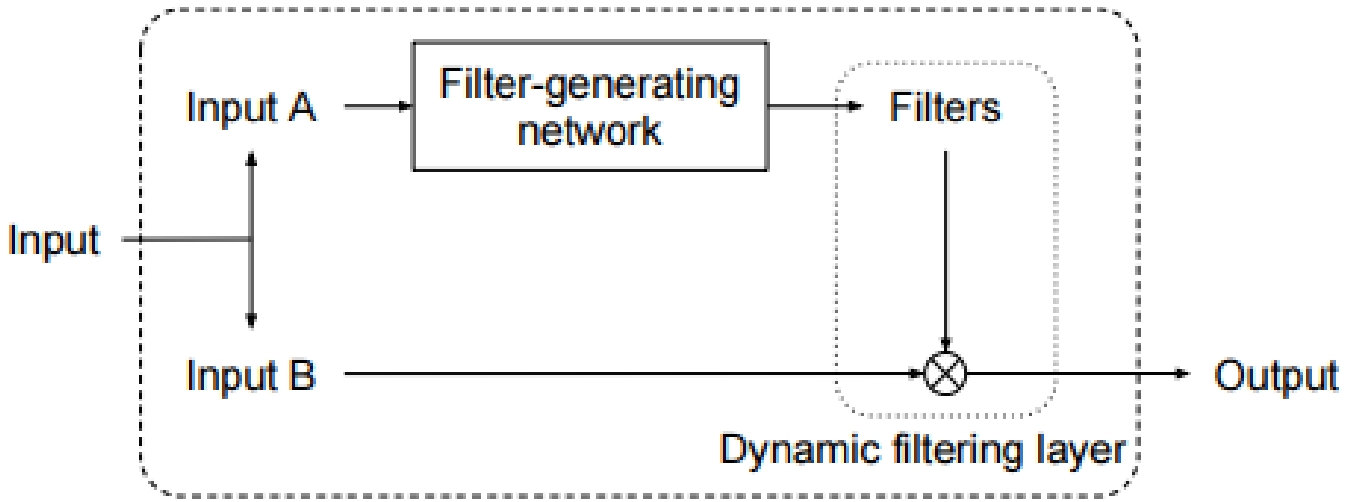
1. Given related views humans can predict the next view
2. Given a video frame humans can predict the next frame
3. Deep Networks trained for the same tasks use the same trained filtering operation for all samples.
4. But, for example in a video frame prediction, different videos may have different motion patterns.
5. Dynamic filters for flexibility : filters conditioned on some input.

Dynamic Filter Network



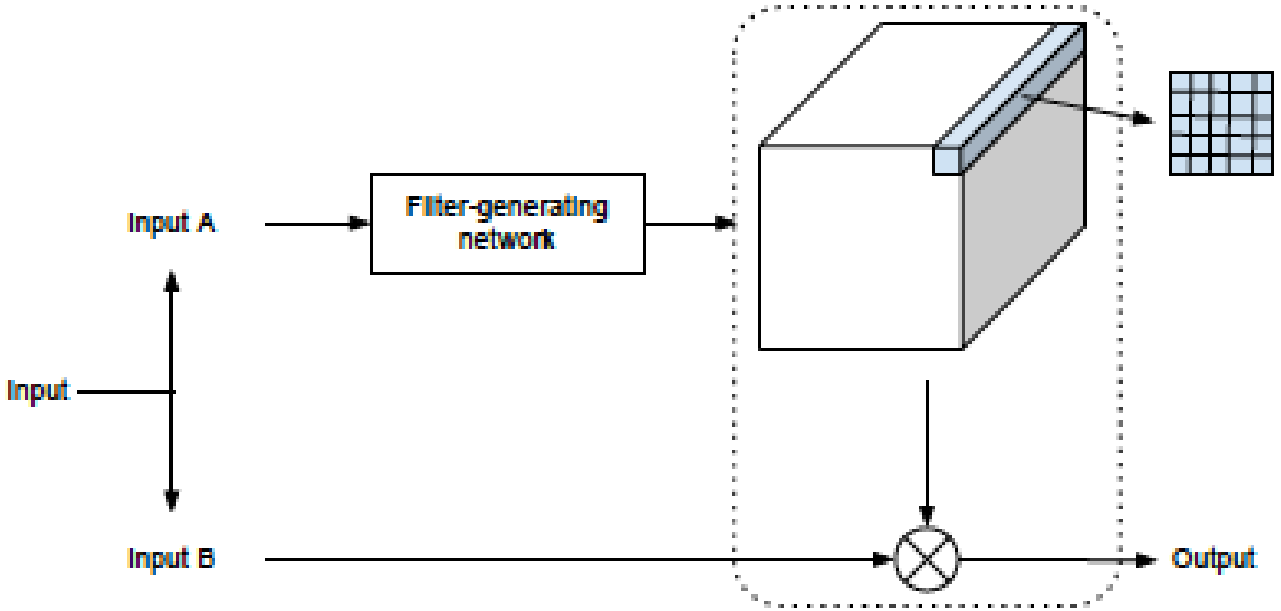
1. *Filter Generating Network* generates filter conditioned on Input A
2. *The Dynamic Filtering Layer* applies it to Input B
3. Input A and B may be the same or different

Filter Generating Network



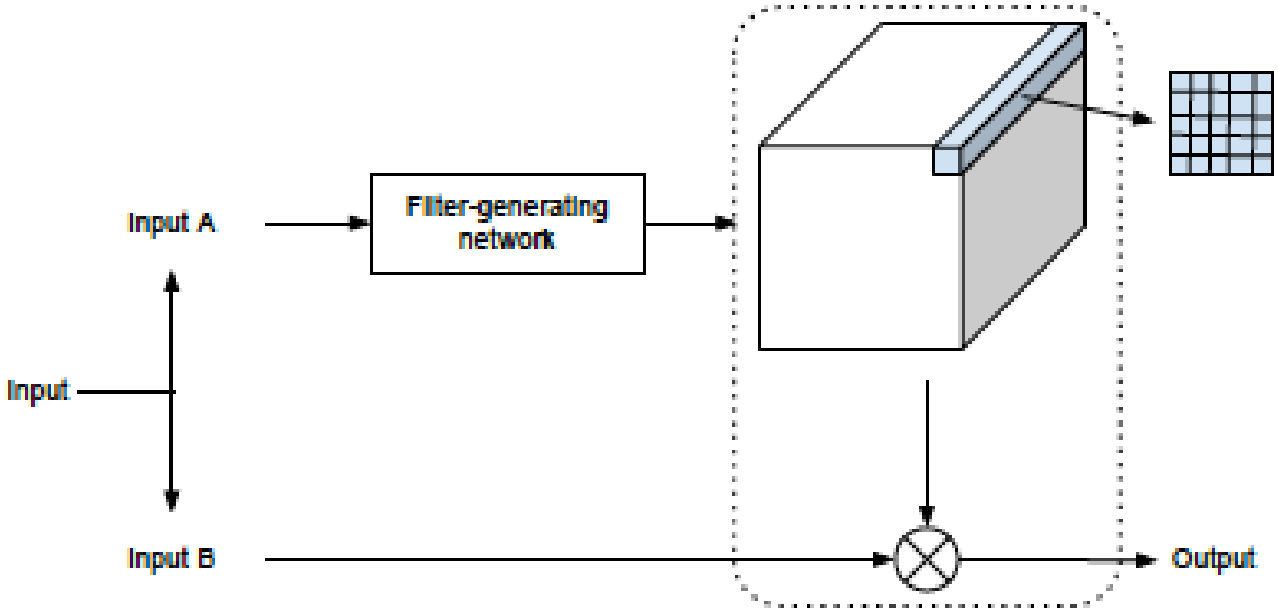
1. Input A $\mathbb{R}^{h \times w \times c_A}$
2. Input B $\mathbb{R}^{h \times w \times c_B}$
3. Generated n Filters F_θ parameterized by $\theta \epsilon \mathbb{R}^{s \times s \times n \times c_B}$ if c_B is the number of channels in Input B

Dynamic Local Filtering Layer



Filters are not same for all spatial locations.
Generated Filters vary for positions on the image
Generated Filters $F_\theta^{(i,j)}$ where $\theta \epsilon \mathbb{R}^{s \times s \times n \times c_B \times h \times w}$ if c_B is the number of channels in Input B
Helps to model local position specific transformation

Dynamic Local Filtering Layer



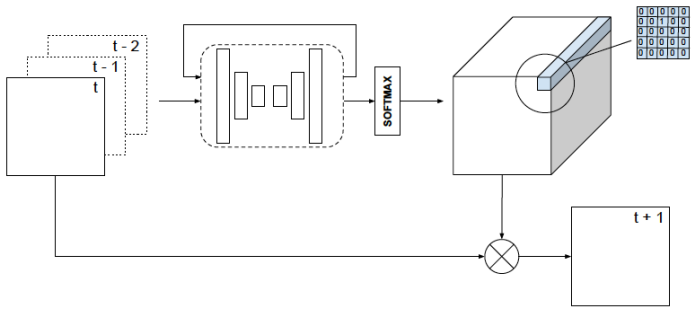
If I_A and I_B are both images, use a CNN for the fiter generating network.

Generated Filters conditioned on corresponding positions in I_A.

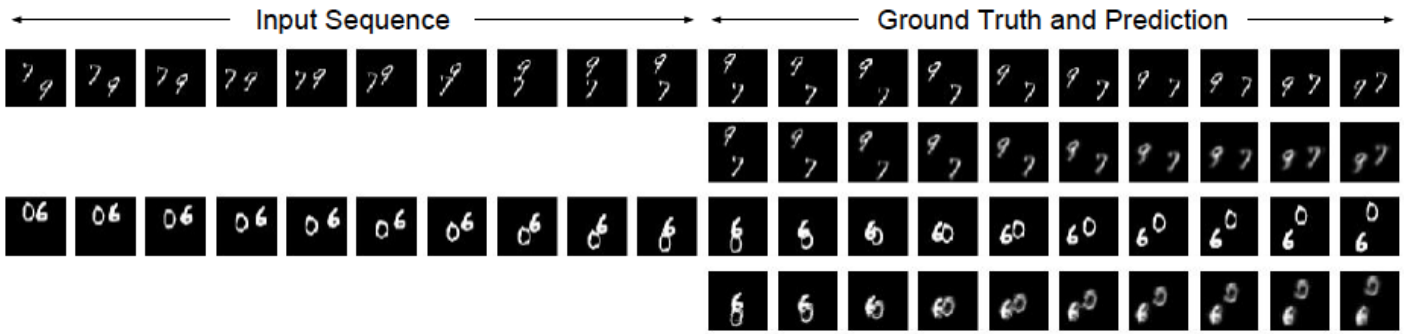
$$G(i,j) = F_\theta^{(i,j)}(I_B(i,j))$$

Results: Video Prediction

1. Task: predict next frame given a sequence of video frames



Model for Video Prediction



Synthetic moving MNIST

Model	Moving MNIST # params	Moving MNIST bce
FC-LSTM [19]	142,667,776	341.2
Conv-LSTM [18]	7,585,296	367.1
Spatio-temporal [15]	1,035,067	179.8
Baseline (ours)	637,443	432.5
DFN (ours)	637,361	285.2

Results: Moving MNIST dataset