

# 文献阅读

20/06/25

# Dynamic Filter Networks

---

**Bert De Brabandere**<sup>1\*</sup>  
ESAT-PSI, KU Leuven

**Xu Jia**<sup>1\*</sup>  
ESAT-PSI, KU Leuven, iMinds

**Tinne Tuytelaars**<sup>1</sup>  
ESAT-PSI, KU Leuven, iMinds

**Luc Van Gool**<sup>1,2</sup>  
ESAT-PSI, KU Leuven  
D-ITET, ETH Zurich

<sup>1</sup>firstname.lastname@esat.kuleuven.be    <sup>2</sup>vangool@vision.ee.ethz.ch

# Introduction

- 有许多工作是围绕着预测展开的：比如视频预测，物体运动轨迹预测等
- 传统做法中的网络经过学习后，会对每一张图片进行完全相同的滤波操作，而这并不是最佳的选择：例如视频预测中，如果先进行运动模式的预测，则可以产生更好的预测效果。
- 本篇工作：利用输入的部分/全部来进行滤波器的生成，以此来根据输入自适应的调整滤波单元

# Method

- 动态滤波方法的核心网络由以下两者组成：
  - 1 Filter-Generating Network 滤波器生成网络
  - 2 Dynamic Filtering Layer 动态滤波层
- 根据需要有两种不同的滤波层：
  - 2.1 Dynamic convolutional layer
  - 2.2 Dynamic local filtering layer

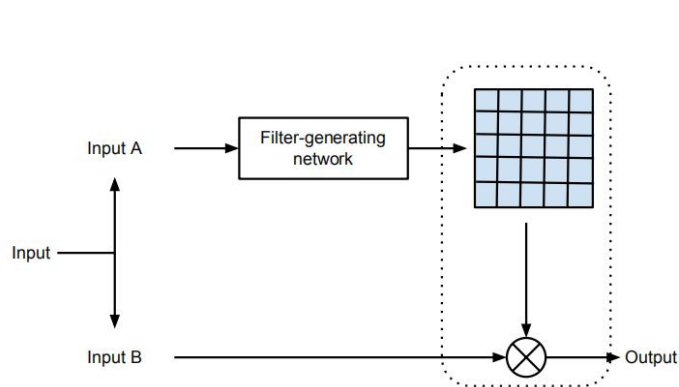
# Method

- 具体参数

若滤波生成网络的输入  $I_A \in \mathbb{R}^{h \times w \times c_A}$

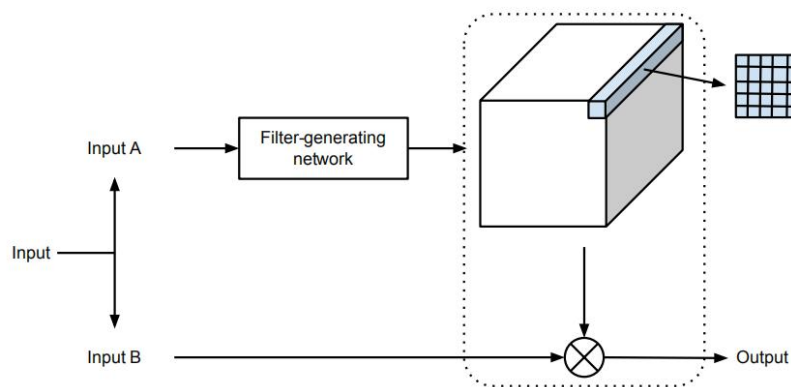
对应的动态滤波层的输入为  $I_B \in \mathbb{R}^{h \times w \times c_B}$

则我们需要生成的参数  $\theta$  为  $\theta \in \mathbb{R}^{s \times s \times c_B \times n \times d}$



(a)

Dynamic convolutional layer



(b)

Dynamic local filtering layer

# Relationship with other networks

- 文献中给出了它与一些经典网络结构的对比：

1 Spatial Transformer Networks

2 Deep Stereo

3 Residual Networks

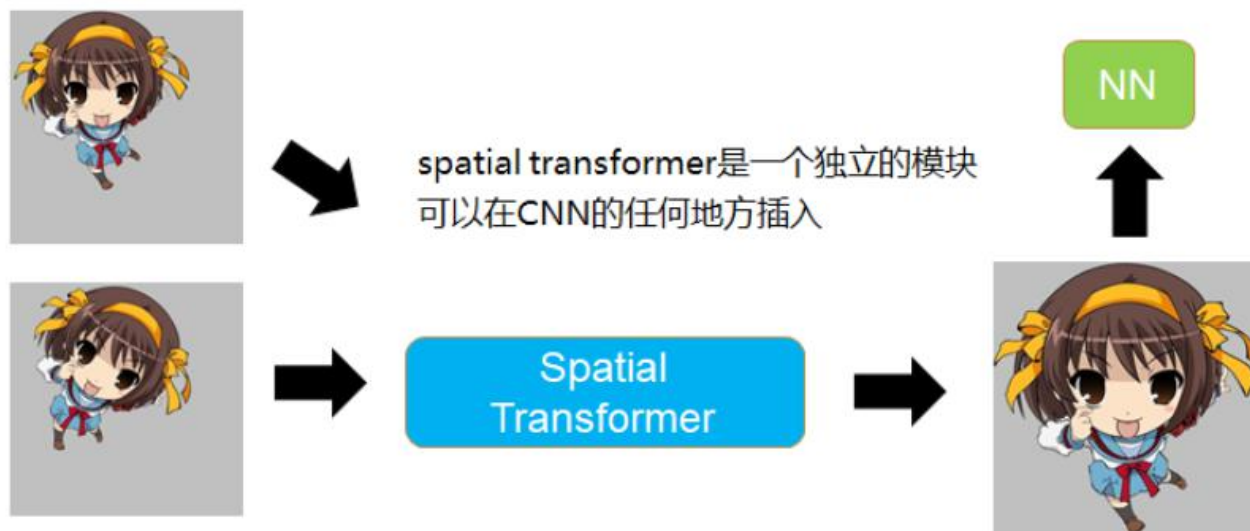
# Relationship with other networks

- Spatial Transformer Networks

原理是先对输入进行一个预处理，将提取到的信息进行系列简单操作如：平移、旋转、裁剪、放缩

相似：网络利用输入来得到一个转换，并应用于特征图

区别：只进行加法乘法运算



# Relationship with other networks

- Residual Networks

原理是双通道，一个通道跳跃连接，另一个通道进行学习，两者相加学习残差

相似：都是双通道，且可以理解为都有一个跳跃连接

区别：残差网络只能两者相加，而动态滤波可以进行更复杂的计算

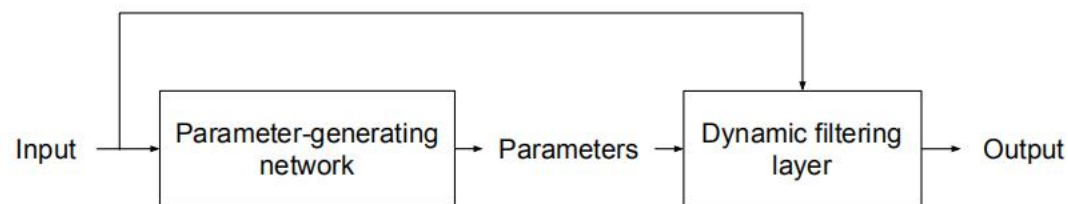
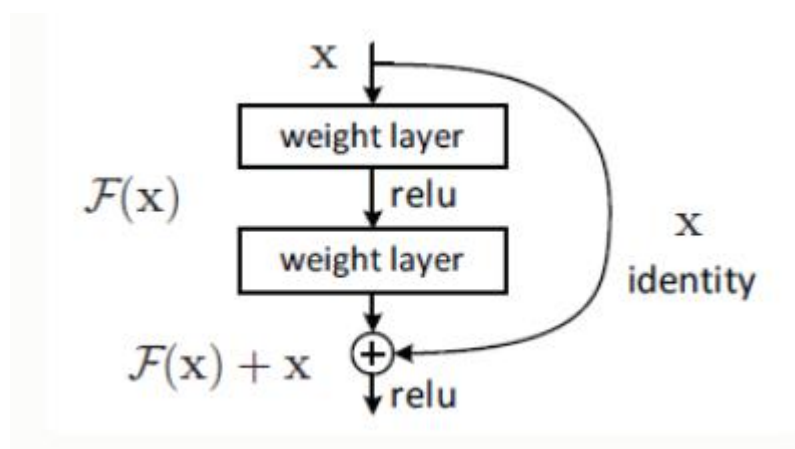


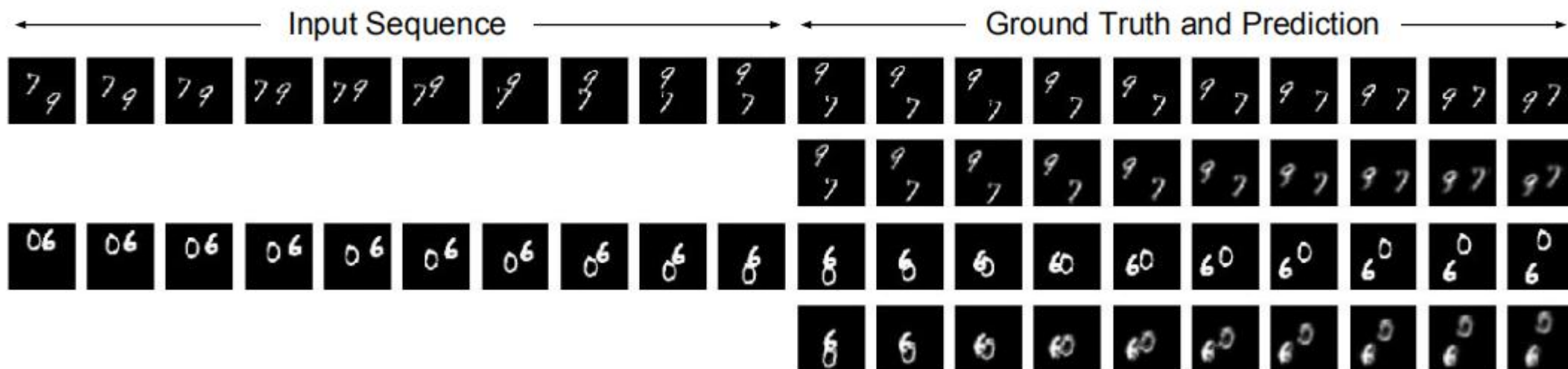
Figure 3: Relation with residual networks.



# Experiments

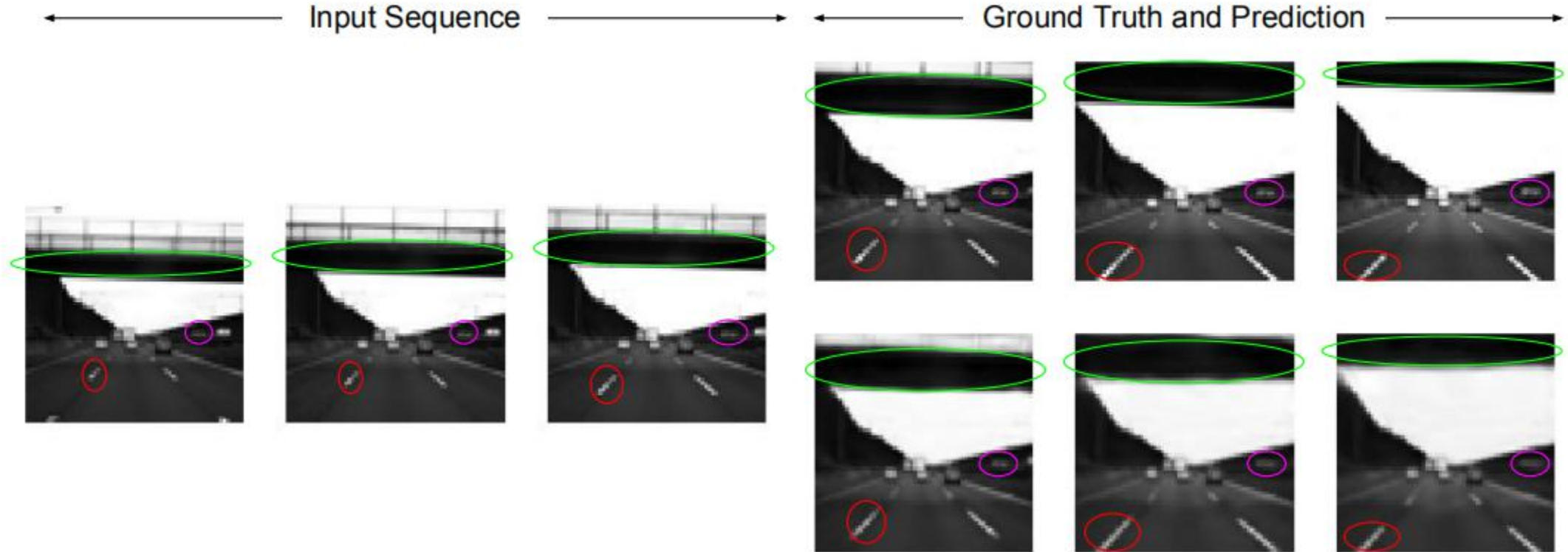
- Dataset: Moving MNIST

Model	Moving MNIST	
	# params	bce
FC-LSTM [21]	142,667,776	341.2
Conv-LSTM [20]	7,585,296	367.1
DFN (ours)	<b>637,361</b>	<b>285.2</b>



# Experiments

- Dataset: Highway Driving



# Method supplement

- 精简版的Dynamic local filtering layer

由于需要生成的参数 $\theta$ 为  $\theta \in \mathbb{R}^{s \times s \times c_B \times n \times d}$

在Dynamic local filtering layer方法中,  $d=h*w$ , 过于庞大, 在文献  
Learning Guided Convolutional Network for Depth Completion

中将其简化了, 即去掉了输入通道数 $c_B$ 这一维度, 使得所有通道使用同一个滤波器(单层, 或称之为卷积核)

然后再利用一个 $1 \times 1$ 的卷积层来学习层间权重。

在一般网络中,  $c_B=32 \sim 128$ , 可以使得网络规模直接缩小这么多倍。

