提出利用神经网络（基于数据学习）中的运动估计和运动补偿来进行视频帧插值。通过CNN估计运动矢量和补偿滤波器。进一步提出了一种基于光流和补偿滤波器的自适应扭曲层来合成新像素。最后使用后处理步骤最小化伪影。

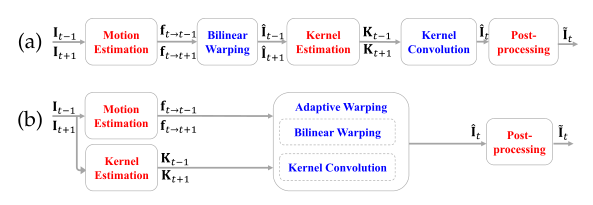
基于学习的方法：**基于光流和基于核的方法**是两种比较好的方法。

（1）其中逐像素（光流）计算运动向量就是一种运动估计，可通过卷积神经网络来估计光流（U-net）。对于运动补偿插值，通常使用重叠块来处理像素块的错误运动向量。

（2）基于核的方法。帧插值可以通过学习自适应局部核通过卷积运算进行插值，不依赖于像素的光流。

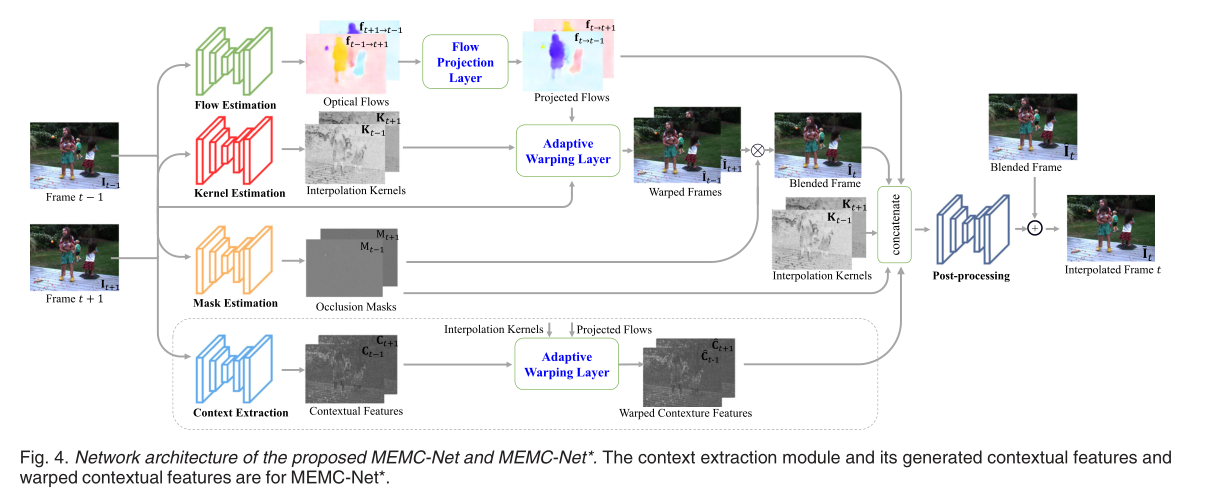
训练数据集通常包含来自原始视频序列的图像三元组，第一帧和第三帧作为输入馈入网络，中间第二帧作为Ground Truth与网络输出得到的像素进行误差计算，模型参数可以通过梯度下降方案迭代更新。

思路框架：

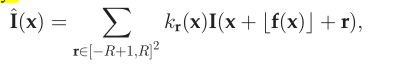


首先对参考帧进行运动估计对齐，对较大的运动采用双线性扭曲处理，对扭曲帧进行小卷积核估计，再进行卷积运算以及后处理步骤合成最终帧。

由于运动估计产生的参考帧质量不好，因此将运动估计（双线性扭曲）和卷积核估计集合在一个步骤里，形成了自适应扭曲层。提出的自适应层基于给定的光流和局部卷积核对图像或特征进行扭曲。



自适应扭曲层算法：



这里 IMG_256 ，前者表示学到的插值核，后者表示双线性参数。

用I(x)表示一个像素从2d坐标到RGB值的映射。

其中x∈[1,H]x[1,W]，表示一个坐标，I(x)表示的就是将坐标映射过去后对应的RGB像素值。

f(x) 表示光流场，其中 u(x), v(x) 分别表示 2d 向量的水平竖直两个部分。



R表示的是卷积核的大小（R\*R）

r∈[-R,R]x[-R,R]，r也表示一个坐标