视频帧插值通常包括两个步骤:运动估计和像素合成。

该方法将插值帧的像素合成看作是两个输入帧的局部卷积。卷积核捕获输入帧之间的局部运动和用于像素合成的系数。

**伪影的产生：**在有遮挡、模糊和亮度突变的区域，光流往往难以估计。基于流的像素合成不能可靠地处理遮挡问题。这两个步骤中的任何一个失败都会导致插入视频帧中出现明显的伪影。

该方法将像素插值看作是对两个输入视频帧中相应图像块的卷积，并利用深度全卷积神经网络估计空间自适应卷积核。

对于在插值帧中（x，y）对应的一个像素，选取了两个参考帧的两个可接受的图像区域R1和R2作为网络的输入，R1和R2都以该像素为中心，最终通过网络估计一个卷积核K。

该卷积核与同样以该像素为中心的图像区域P1和P2进行卷积运算得到该像素对应的中间帧像素。

将视频所有帧分成三帧组，每一帧包含一个视频中的三个连续帧。然后在每个三帧组中随机选取一个像素，并从视频帧中提取以该像素为中心的三patch组。