Question

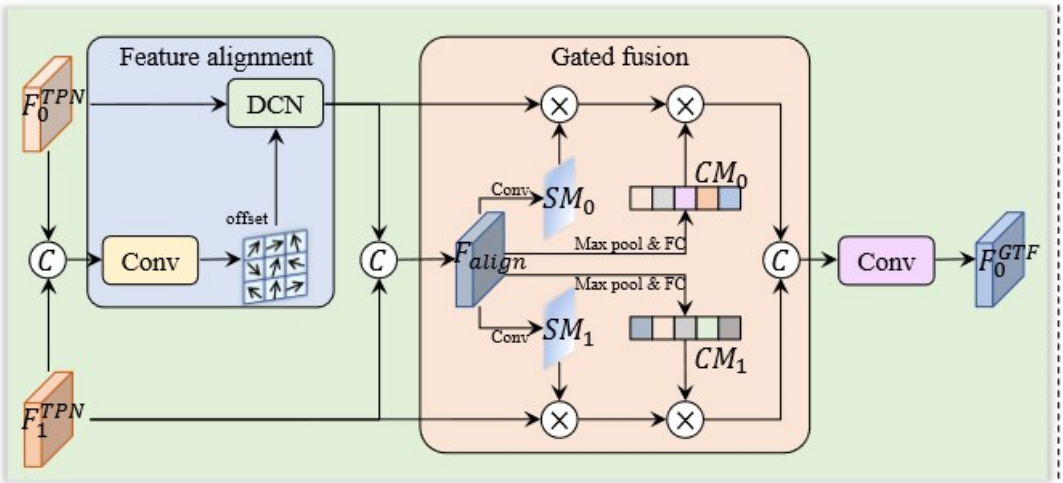
1. RecEvSR

1.1 评估Metric RMSE是对SR ECM计算？还是SR Event stream计算?

如果是对SR ECM计算，不就意味着在训练、评估时都忽略了从SR ECM恢复到SR Event stream的temporal误差吗？但是Event SR任务的目标不是基于LR Event stream超分得到HR Event stream。

1.2 为什么Event SR任务不像Reconstruction、VFI任务，选择voxel作为Event representation，而是ECM？不会损失很多temporal information吗?

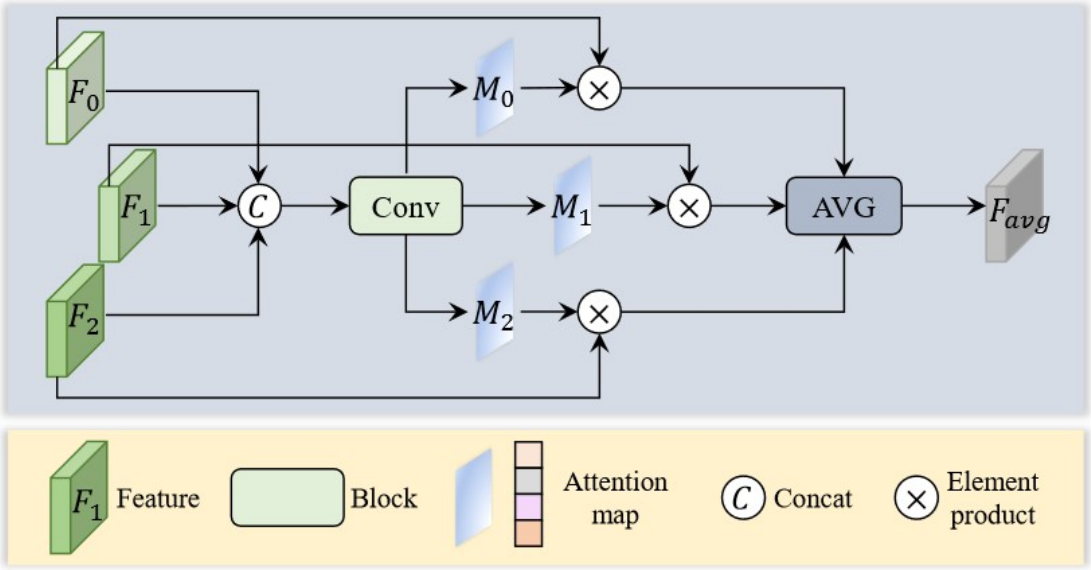
1.3 为什么不显示使用额外的loss约束，就可以确保DCN能将F_1 align到F_2的timestamp呢?



1.4 为什么如此设计ASTF的结构?

ASTF在fusion同一scale下不同timestamp特征前，先计算attention建模temporal context。

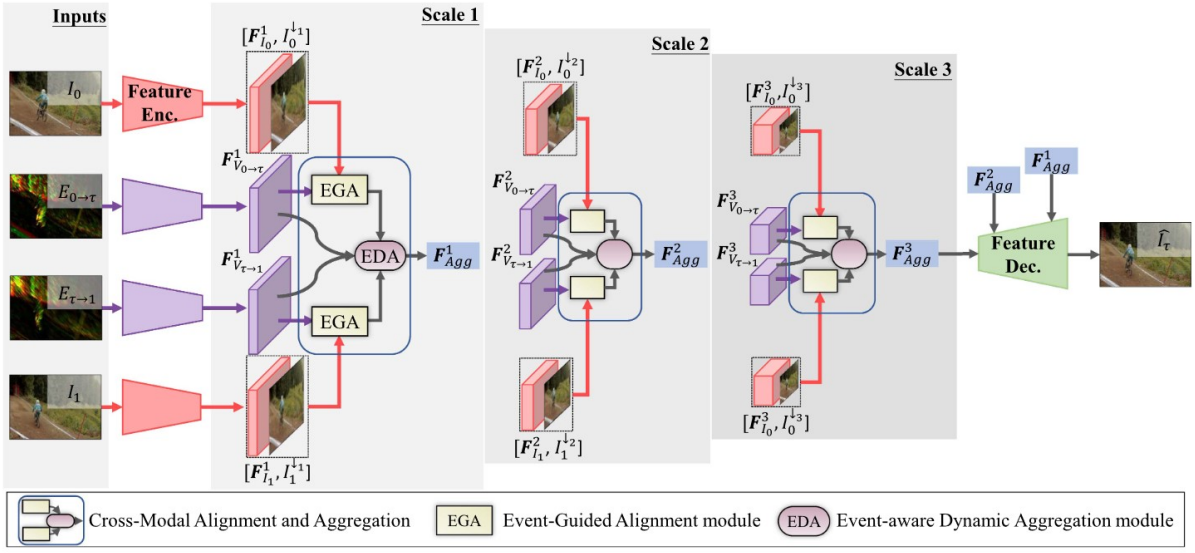
通常，计算attention是以当前时间戳特征作为Q，所有时间戳特征作为K,V。为什么这里是反过来，以其它时间戳特征为Q，当前时间戳特征作为K,V，这样设计的目的是什么呢?



2. EVA2

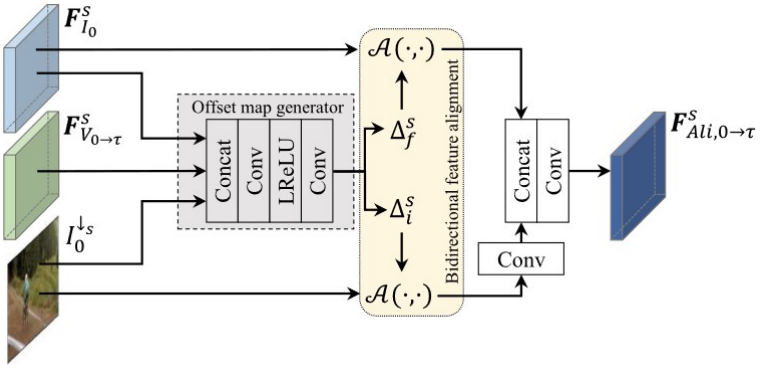
2.1 为什么在align得到intermediate outputs后，还需要再额外aggregate event data来提升性能?

Align和Aggregate本质上都是在Fusion Event data和RGB data。如果说Aggregate event data的目的是利用Event data消除corruption，那为什么在Align时fusion event data时没有这个效果呢?



2.2 为什么concat image效果这么好，这是VFI的特点吗？还是low-level vision的特点?

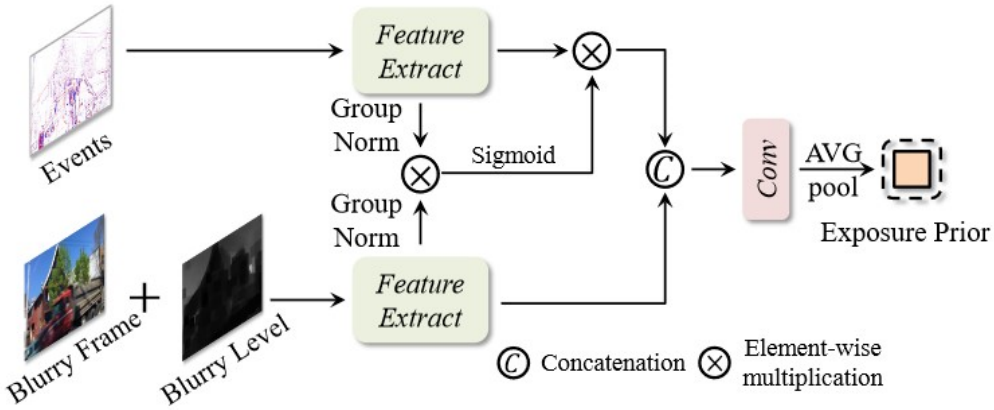
encoder提取各scale特征，还不如直接下采样image有用？感觉很不符合常理。



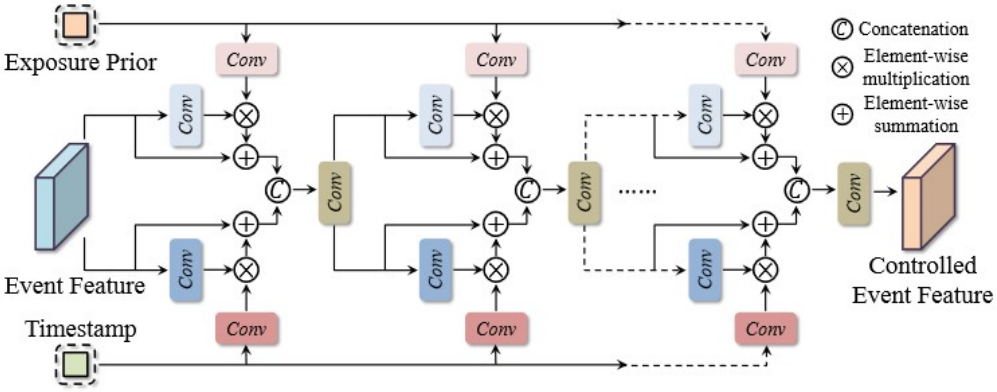
Method	#Params. (M)	PSNR
EGA *-w/o Event*	3.68	35.98
EGA *-w/o Image*	3.68	36.30
EGA *-w/o Feature*	3.67	36.41
EGA - *wider*	3.90	36.63
EGA - *deeper*	3.90	36.65
EGA - *rescale*	3.69	36.15
EGA	3.69	36.51

3.Event-based Blurry Frame Interpolation under Blind Exposure

3.1 作为一个end2end系统，怎么保证Exposure prior就是T_{ex}/ (T_{ex} + T_{re})?



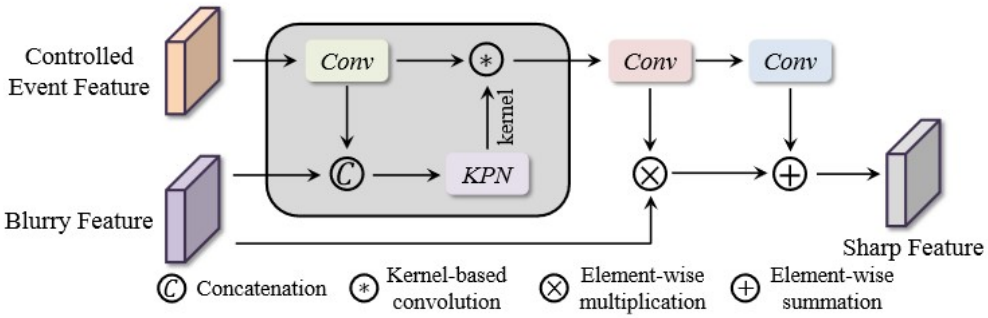
3.2 为什么需要做成多步骤？每步的controlling coefficients是一样的，也没有改变scale



3.3 KPN就是Dynamic Conv?

3.4 为什么最后执行相乘、相加操作?

相乘是单纯为了fusion两个特征吗？既然都已经是feature，与完全EDI不一样了吧。



3.5 为什么不考虑更大范围的Blurry frame和Event?

---

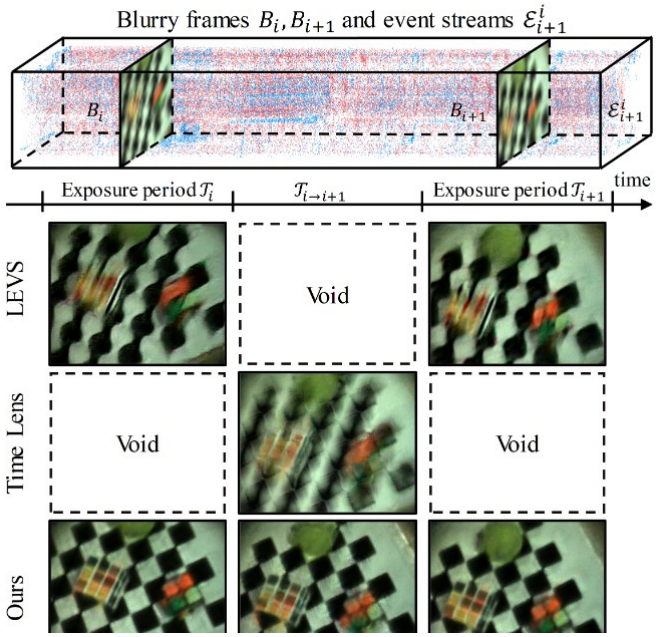
4. Conclusion

a）Standard camera会损失exposure期间的信息，导致帧率低。且如果exposure期间相机和物体存在相对移动，还会导致blurry

b）引入Event data改善这一点，理想情况 **i) Event无噪声干扰，ii) threshold是常数**:

根据EDI可以得到任意timestamp的latent sharp frame，凭此可以提升大多数frame-based task效果(VSR)

4.1 video的exposure time外为什么还有间隔?



为了近似这一理想状况：

a）降低Event噪声干扰：Representation大多采用window of event

b）threshold未知：建模threshold的时空分布

a）显式地align到 latent sharp frame，隐式建模地threshold

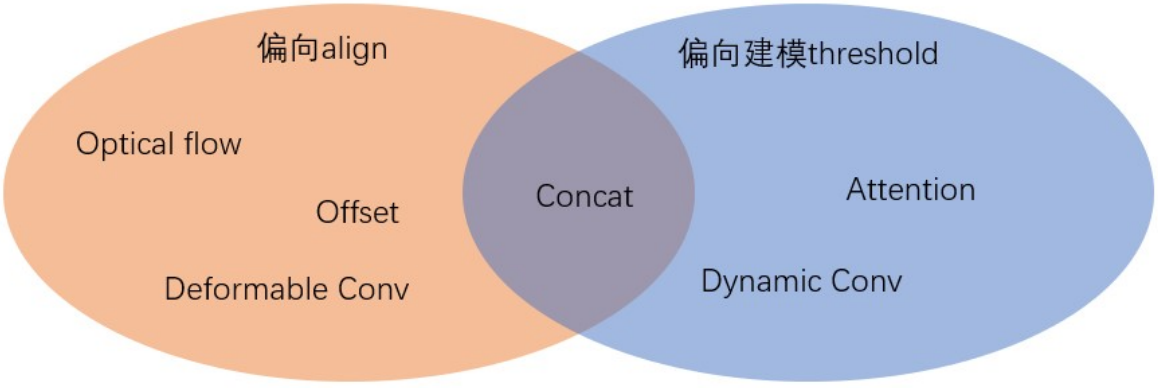
4.2 为什么在Turning Frequency to Resolution，Dynamic Conv可以直接用于align?

Specifically, given a set of successive LR frames $\{\mathcal{I}_{t_i}^{LR}\}$ and also the uniform event representations $\{E_u\}_{[t_{i-1},t_i]}^{LR}$ in the corresponding time intervals, the proposed LR-BAI loss can then be formulated as:

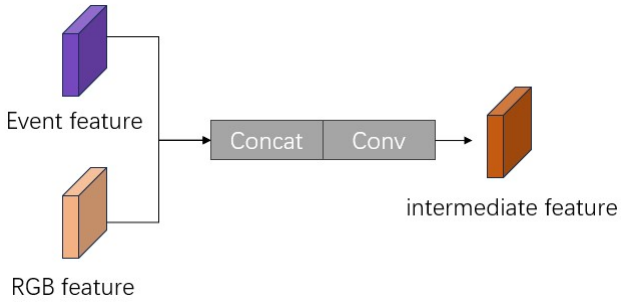
$$\mathcal{L}_{BAI}^{LR} = \sum_i \Big(\Big\| f_{\theta^*}\left(\mathcal{I}_{t_{i-1}}^{LR}, \{E_u\}_{[t_{i-1},t_i]}^{LR}\right) - \mathcal{I}_{t_i}^{LR} \Big\| + \Big\| f_{\theta^*}^{-1}\left(\mathcal{I}_{t_i}^{LR}, \{E_u\}_{[t_{i-1},t_i]}^{LR}\right) - \mathcal{I}_{t_{i-1}}^{LR} \Big\| \Big), \tag{5}$$

where f_{θ^*} is the dynamic convolutional operation. According to the property of event data, with the optimal adaptively generated parameters θ^* from $\{E_u\}_{[t_{i-1},t_i]}^{LR}$, the LR frame $\mathcal{I}_{t_{i-1}}^{LR}$ at a specific timestamp can be convolved into the frame $\mathcal{I}_{t_i}^{LR}$ at the next timestamp, which inspires the design of the proposed loss function. Furthermore, we impose a bi-directional interpolation constraint in Eq. 5, where the aforementioned transformation should be conducted in both directions via a pair of dynamic convolution f_{θ^*} and reverse dynamic convolution $f_{\theta^*}^{-1}$, as shown in Fig. 5.

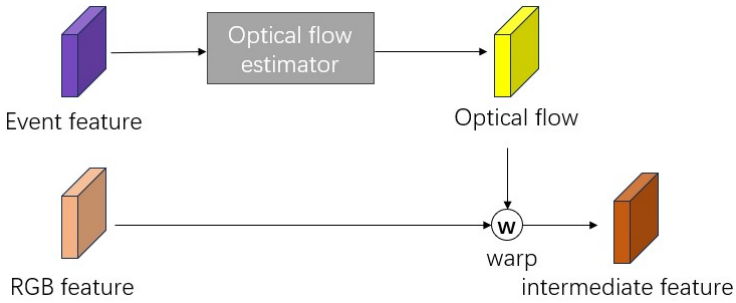
b）显式地建模threshold，隐式地align



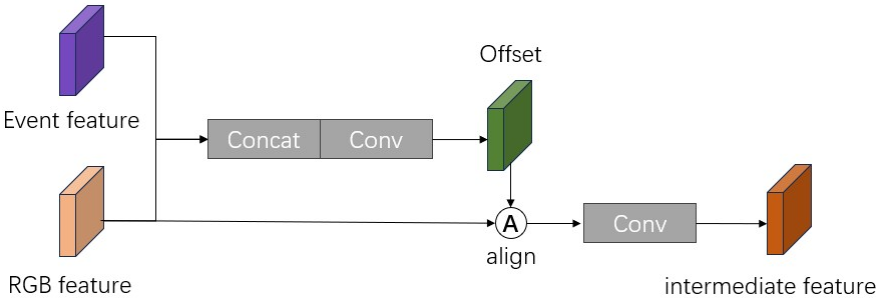
a）Concat



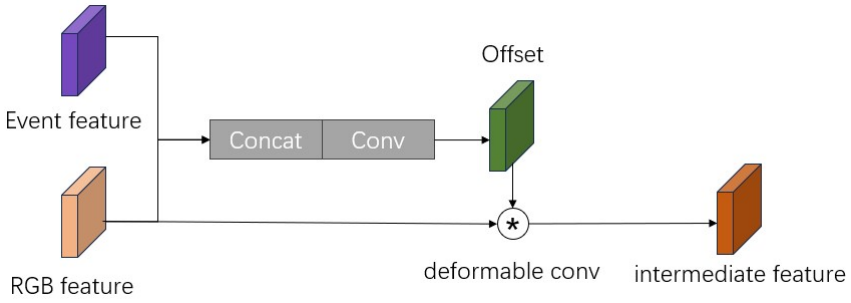
b) Optical Flow



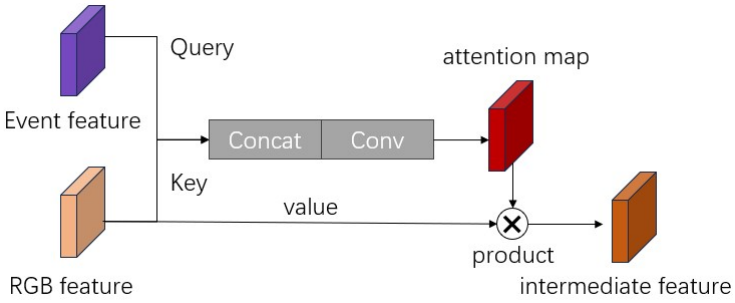
c) Offset



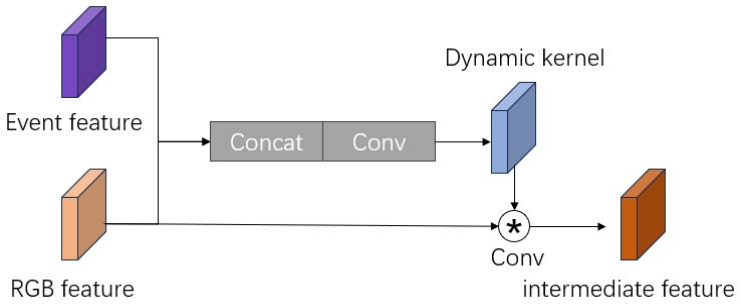
d) Deformable Conv



e) Attention



f) Dynamic Conv



4.3 c) Offset 和 b) Optical flow 的区别是什么？EVA2中对区别的解释没看懂

4.4 可以解释一下为什么Dynamic Conv > Concat > attention吗?

- 因为Dynamic kernel比concat更显式地对threshold进行建模，所以效果更好?
- 为什么Concat比attention要好？（attention比concat更显式建模threshold）
- 为什么Dynamic Conv比attention要好？（除了product和conv操作不一样，同样都对每个pixel建模了threshold）

Method	#Params. (M)	PSNR
EDA - *Concat*	3.69	36.37
EDA - *Attention*	3.60	36.28
EDA - *DynamicConv*	3.68	36.39
EDA - *wider*	3.78	36.58
EDA - *deeper*	3.78	36.60
EDA	3.69	36.51

4.5 输入Dynamic Conv，channel恢复到Bin，显式对每个Bin建模threshold是不是更好?

更符合threshold对时空都变化的特点

4.5 为什么后续不怎么采用transformer结构了?