11.14-11.20

EVEN: An Event-Based Framework for Monocular Depth Estimation at Adverse Night Conditions

Peilun Shi1, Jiachuan Peng1, Jianing Qiu1,2,†, Xinwei Ju1, Frank Po Wen Lo1, and Benny Lo1

Method

事件数据可以捕获更多的 HDR 和夜间场景的时间细节，而 RGB 数据可以提供必要的纹理和颜色信息。`10.24` 这一篇是说帧帧数据可以在几乎没有自我运动或很少触发事件的情况下提供重要的参考信息。这是因为事件数据通常只捕捉到亮度变化，而帧数据可以提供更多关于场景的详细信息，有助于更好地理解和处理那些相对静止的情境

使用基于事件的视觉和低光图像增强，将增强后的RGB图像与事件图像融合，重建图像，估计不利夜间条件下的单眼深度的工作。

A.Event Stream（处理事件数据）

将体素网格格式的事件流转换为图像格式。沿时间轴 t 使用固定时间段 Δt = 0.125 s 堆叠产生一个紧凑的事件图像。（简单的按时间堆叠）

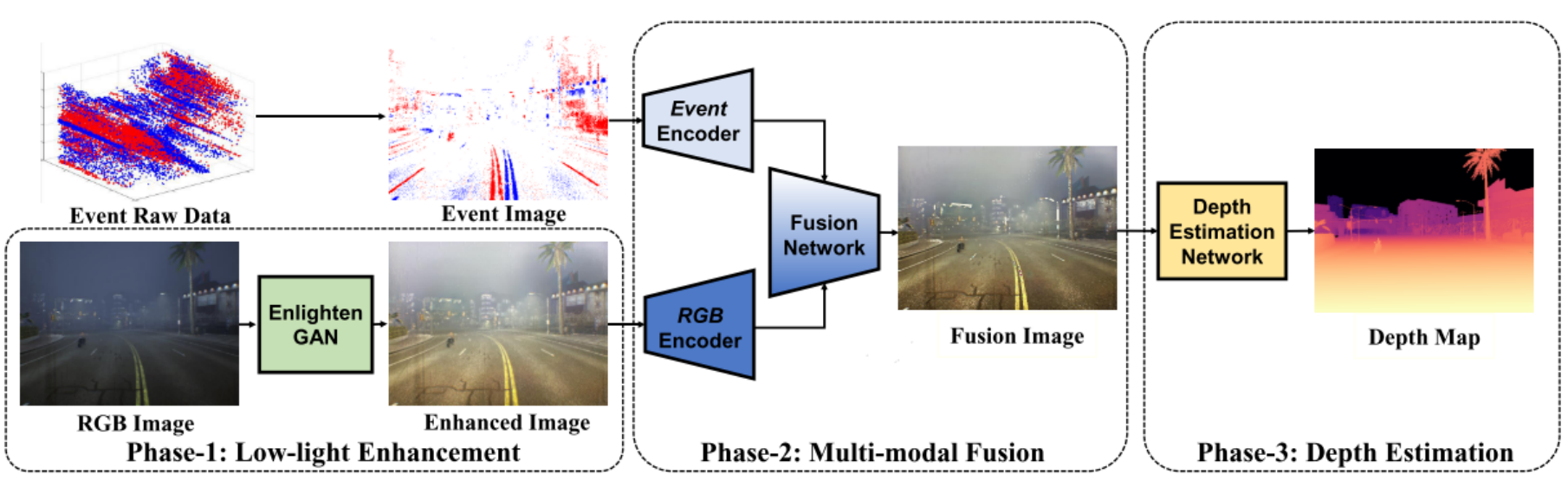
B. Phase-1: Low-light Enhancement

利用 EnlightenGAN来增强原始夜间 RGB 图像。 EnlightenGAN 是一种基于注意力的 U-Net 结构。输入是原始的夜间RGB图像。在处理中，RGB图像被规范化，并使用亮度通道作为注意力图。这个注意力图指导增强过程，使网络能够更专注于图像中的关键区域。

> 亮度通道是指图像中表示亮度（明亮程度）的部分。通过将亮度通道作为注意力图，模型可以更加关注图像中的亮部或暗部，从而更有针对性地进行图像增强。亮度通道可以通过对RGB通道进行适当的加权和组合来得到，例如：$Y=0.299\cdot R+0.587\cdot G+0.114\cdot B$*（常见）*

C. Phase-2: Multi-modal Fusion

为了充分利用它们的优点，作者设计了一个新颖的融合网络，基于选择性核网络构建，用于集成事件数据和RGB模态。



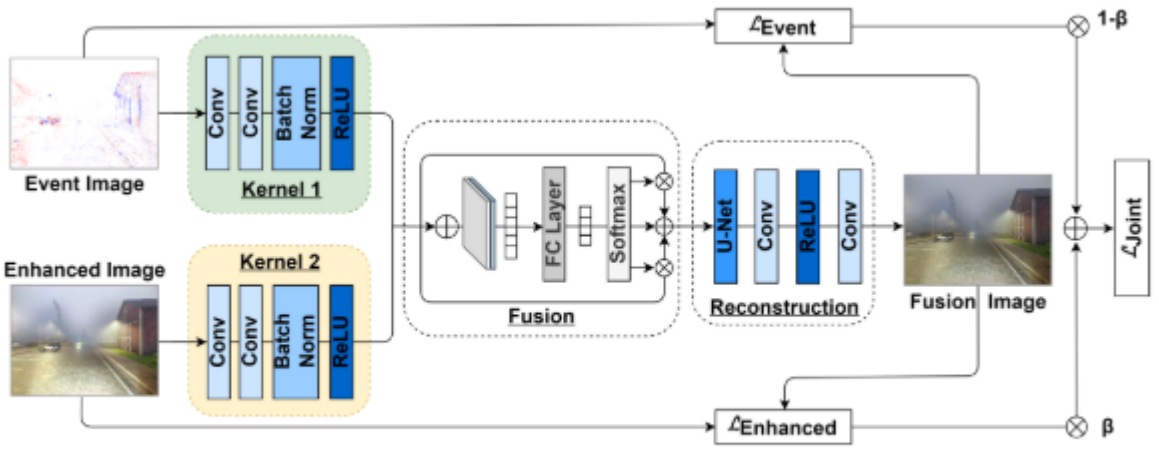


Fig. 3: The multi-modal fusion network of EVEN.

（1）Fusion Network

1. **输入：** 该网络接受两个输入图像，一个是事件图像 **X***Event*，另一个是增强的RGB图**X***Enhanced*。
2. **特征提取：** 通过使用具有不同核大小的两个卷积核，将输入图像转换为特征图。分别用 *g*(·) 和 *h*(·)表示这两个转换层。对于事件图像，采用 5×5×5 的大核，因为事件模态携带的信息相对稀疏。对于增强的RGB图像，采用 3×3×3 的小核。
3. **特征融合：** 将两个模态的特征图进行元素级求和，得到融合后的特征图 **F***sum*。

$$\mathbf{F}_{Event} = g(\mathbf{X}_{Event}), \mathbf{F}_{Event} \in \mathbb{R}^{H \times W \times C}$$

$$\mathbf{F}_{Enhanced} = h(\mathbf{X}_{Enhanced}), \mathbf{F}_{Enhanced} \in \mathbb{R}^{H \times W \times C}$$

$$\mathbf{F}_{sum} = \mathbf{F}_{Event} + \mathbf{F}_{Enhanced}, \mathbf{F}_{sum} \in \mathbb{R}^{H \times W \times C}$$

4. **全局平均池化：** 对融合后的特征图 **F***sum* 进行全局平均池化，降低特征图的维度，得到一个紧凑的向量 **V(**V ∈ R1×C)
5. **全连接层：** 使用一个简单的全连接层 *f*(·) 将向量**V** 转换为一个紧凑的向量 **k**，其维度为 *d*×1。

$$\mathbf{k} = f(\mathbf{V}), \mathbf{k} \in \mathbb{R}^{d \times 1}$$

6.**自适应融合：** 利用向量 **k** 对两个模态进行自适应融合，通过学习的权重 *ac* 和 *bc* 对事件和增强的RGB特征图进行加权融合。

$$a_c = \frac{e^{\mathbf{A}_c \mathbf{k}}}{e^{\mathbf{A}_c \mathbf{k}} + e^{\mathbf{B}_c \mathbf{k}}}, b_c = \frac{e^{\mathbf{B}_c \mathbf{k}}}{e^{\mathbf{A}_c \mathbf{k}} + e^{\mathbf{B}_c \mathbf{k}}}$$

$$\mathbf{F}_{fused_c} = a_c \cdot \mathbf{F}_{Event_c} + b_c \cdot \mathbf{F}_{Enhanced_c}, a_c + b_c = 1$$

7.**UNet和重构：** 将融合后的特征图输入到UNet中，随后通过一系列的卷积和ReLU操作进一步融合事件和RGB模态的特征，并重构出具有相同分辨率的融合图像 **Y**。

$$\mathbf{Y} = \mathrm{Conv}(\mathrm{ReLU}(\mathrm{Conv}(\mathrm{U-Net}(\mathbf{F}_{fused}))))$$

(2) Fusion Loss

1. **联合损失设计：** 作者设计了一个联合损失函数 L*joint*，该损失函数由主要损失项 L*Enhanced* 和辅助损失项 L*Event* 组成。
2. **主要损失：** 主要损失L*Enhanced* 是融合图像与增强的RGB图像之间的重建损失。它通过计算两者之间的均方误差（L2损失）来衡量，表示为图像之间的平均差异。
3. **辅助损失：** 辅助损失L*Event* 是融合图像与事件图像之间的重建损失。同样，它也通过计算均方误差来度量两者之间的差异。
4. **损失加权：** 联合损失L*joint* 通过加权组合主要损失和辅助损失。权重参数 *β* 调节了主要损失在整体损失中的相对贡献。
5. **训练目标：** 在训练过程中，融合网络的目标是减小联合损失L*joint*，通过调整网络参数使融合图像更好地逼近增强的RGB图像和事件图像。

$$\mathcal{L}_{joint} = \beta \times \mathcal{L}_{Enhanced} + (1 - \beta) \times \mathcal{L}_{Event}$$

D. Phase-3: Depth Estimation

1. **融合图像作为输入：** 融合网络生成的融合图像包含了来自事件和RGB两种模态的视觉信息。这个融合图像被用作深度估计的输入。
2. **深度估计网络：** 作者采用了两种先进的深度估计网络，分别是 Depthformer 和 SimIPU 。这两个网络被用于对融合图像进行深度估计。
3. **Depthformer：** Depthformer 是一种深度估计网络，用于从图像中推断出场景的深度信息。
4. **SimIPU：** SimIPU 也是一种深度估计网络，用于对图像进行深度推断。

DATASET

构建了第一个不良夜间驾驶数据集

MonoANC(Monocular depth estimation at Adverse Night Conditions).

EXPERIMENT

Baseline Methods

介绍了六种基准方法，用于比较和检验作者提出的框架对于提升深度估计的效果。这些基准方法主要包括：

1. **RGB：** 将原始RGB图像直接输入深度估计网络，作为深度估计的唯一输入。
2. **Event：** 将事件图像直接输入深度估计网络，作为深度估计的唯一输入。
3. **RGB + Sobel：** 将原始RGB图像和经过Sobel算子处理的图像一起作为输入，经过EVEN的第二阶段，然后进行深度估计。
4. **RGB + Event：** 将原始RGB图像和事件图像一起作为输入，经过EVEN的第二阶段，然后进行深度估计。
5. **RGBEnhanced：** 将经过增强的RGB图像（在第一阶段之后）直接输入深度估计网络，作为深度估计的唯一输入。
6. **RGBEnhanced + Sobel：** 将经过增强的RGB图像和经过Sobel算子处理的图像一起作为输入，经过EVEN的第二阶段，然后进行深度估计。

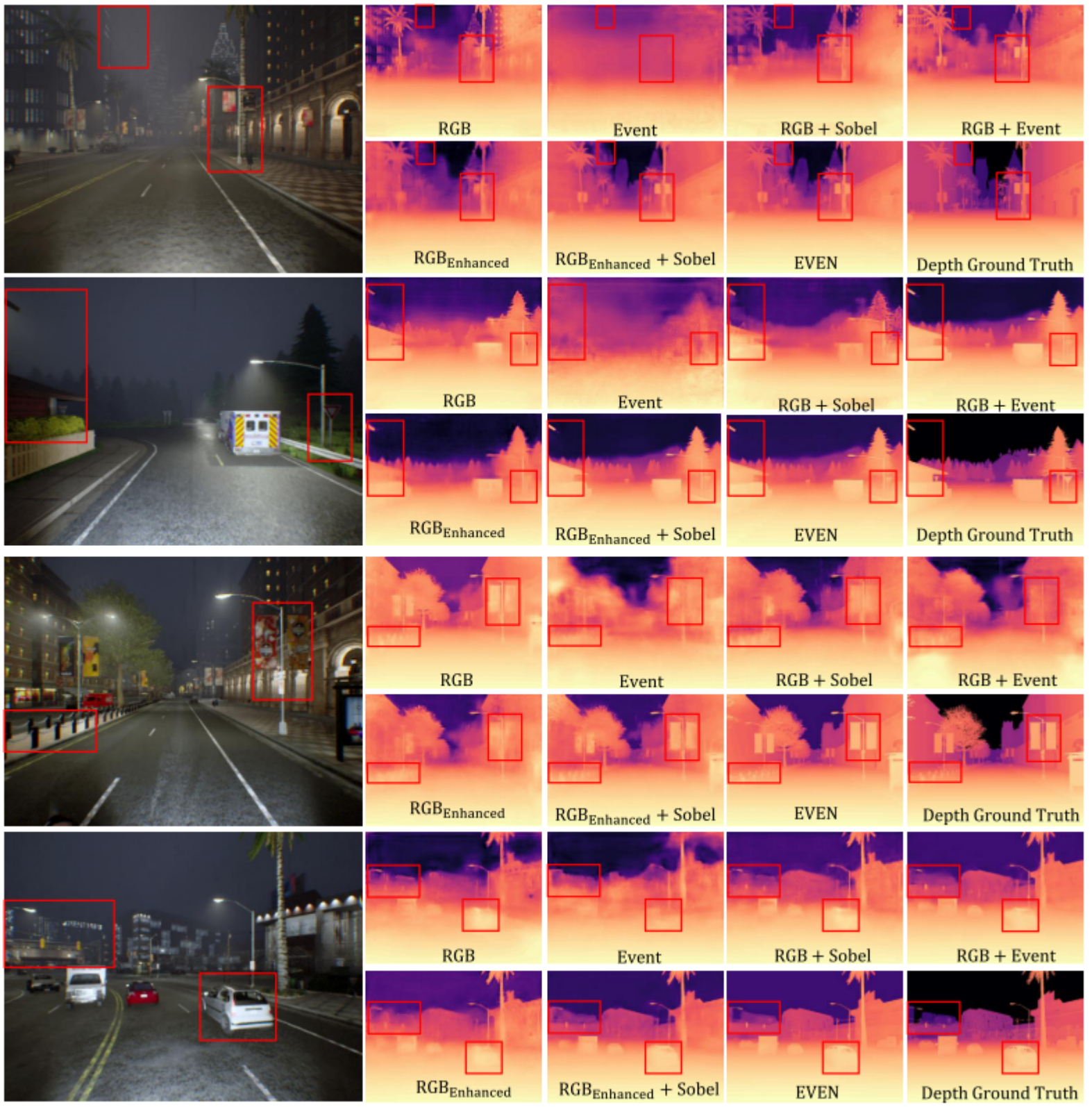
这些基准方法的设计旨在探索和验证提出框架在深度估计任务上的有效性，尤其是通过低光环境下的图像增强和事件与RGB模态的融合。

Overall Results

EVEN框架在深度估计任务中的表现优于基准方法，且在不同深度估计网络（Depthformer和SimIPU）上表现一致。

Input Sequence	Depthformer							SimIPU						
	Error Metric ↓				Accuracy Metric ↑			Error Metric ↓				Accuracy Metric ↑		
	Abs. Rel.	Sq. Rel.	RMSE	Log10	$\alpha 1$	$\alpha 2$	$\alpha 3$	Abs. Rel.	Sq. Rel.	RMSE	Log10	$\alpha 1$	$\alpha 2$	$\alpha 3$
RGB	0.192	0.310	4.973	0.069	0.810	0.911	0.985	0.293	0.370	5.177	0.079	0.710	0.921	0.972
Event	0.452	0.220	7.775	0.172	0.390	0.622	0.795	0.594	1.240	9.180	0.116	0.552	0.828	0.932
RGB + Sobel	0.180	0.340	5.304	0.064	0.808	0.908	0.956	0.266	0.310	4.947	0.067	0.773	0.930	0.976
RGB + Event	0.179	0.340	5.992	0.067	0.795	0.920	0.956	0.229	0.280	5.151	0.057	0.837	0.953	0.984
$\text{RGB}_{Enhanced}$	0.181	0.390	5.737	0.074	0.765	0.924	0.971	0.263	0.300	4.998	0.058	0.824	0.948	0.984
$\text{RGB}_{Enhanced}$ + Sobel	0.139	**0.280**	5.023	0.063	0.806	0.970	0.988	0.216	**0.240**	**4.080**	0.063	0.846	0.954	0.986
EVEN (Ours)	**0.112**	**0.280**	**4.335**	**0.049**	**0.903**	**0.976**	**0.993**	**0.125**	0.280	4.845	**0.049**	**0.857**	**0.959**	**0.988**

EVEN中的深度估计网络分别实例化为Depthformer和SimIPU时在MonoANC数据集上的结果



Cross Validation on Adverse Weather

根据不同的天气情况进一步拆分了 MonoANC。具体来说，如图4（b）所示，存在三种不利天气条件：1）仅下雨； 2) 仅雾； 3）雨和雾同时发生。

当深度估计框架在训练过程中接触每种单独的天气条件时，能够很好地估计混合恶劣天气条件的场景，即雨和雾同时发生。然而，当模型在训练中接触到混合天气条件的样本时，难以对单一天气条件的进行深度估计。因此，作者提到在未来的工作中值得研究的是，通过拆分恶劣天气组合的代价函数来提高深度估计的性能。

Input Sequence		Depthformer							SimIPU						
		Error Metric ↓				Accuracy Metric ↑			Error Metric ↓				Accuracy Metric ↑		
Train Set	Test Set	Abs. Rel.	Sq. Rel.	RMSE	Log10	$\alpha 1$	$\alpha 2$	$\alpha 3$	Abs. Rel.	Sq. Rel.	RMSE	Log10	$\alpha 1$	$\alpha 2$	$\alpha 3$
ain and fog at the same time	rain only and fog only	0.325	1.987	8.475	0.187	0.471	0.645	0.797	0.330	1.865	8.710	0.187	0.420	0.655	0.786
rain only and fog only	rain and fog at the same time	**0.267**	**0.315**	**4.934**	**0.031**	**0.646**	**0.833**	**0.937**	**0.260**	**0.307**	**4.933**	**0.031**	**0.680**	**0.844**	**0.939**

EVEN 在不同恶劣天气条件下的交叉验证结果

CONCLUSION

本文提出了一种框架，通过集成低光增强和融合RGB和事件模态，实现在恶劣的夜间条件下有效的单目深度估计。构建了一个包含配对的RGB、事件和深度图像的合成夜间驾驶数据集，其中包括在恶劣天气、光照和道路条件下的场景。实验证明，提出的框架能够在各种恶劣的夜间情景中实现令人满意的深度估计结果。