LiDAR-Event Stereo Fusion with Hallucinations (ECCV 2024)

🚨 This repository contains download links to our code, and trained deep stereo models of our work "**LiDAR-Event Stereo Fusion with Hallucinations**", ECCV 2024

by Luca Bartolomei[1,2], Matteo Poggi[2], Andrea Conti[2], and Stefano Mattoccia[1,2]

Advanced Research Center on Electronic System (ARCES)[1] University of Bologna[2]

LiDAR-Event Stereo Fusion with Hallucinations (ECCV 2024)

Project Page | Paper | Supplementary

Note: 🚧 Kindly note that this repository is currently in the development phase. We are actively working to add and refine features and documentation. We apologize for any inconvenience caused by incomplete or missing elements and appreciate your patience as we work towards completion.

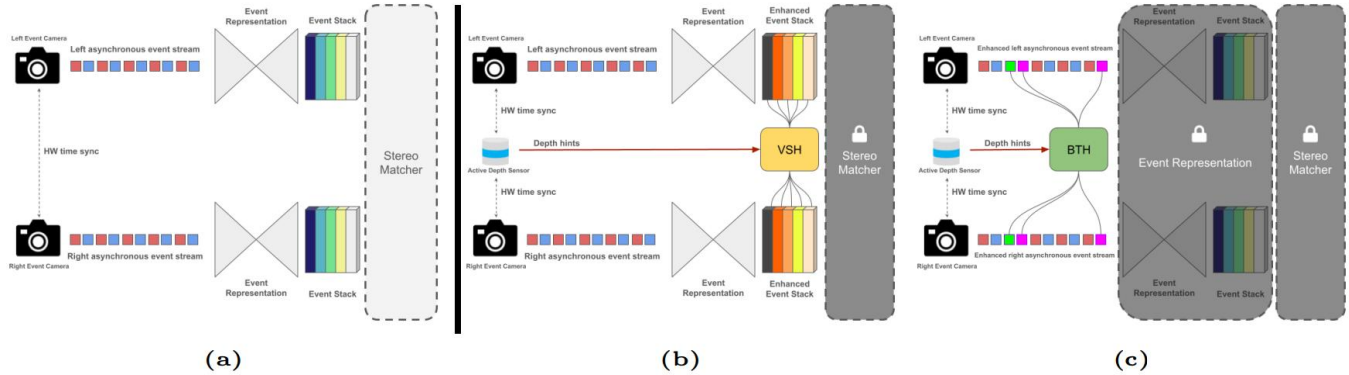
Table of Contents

Introduction

This pioneering paper proposes a solution to the issues caused by the absence of motion or the presence of large untextured regions in an event-based stereo matching setup.

Given that event cameras provide rich cues at object boundaries and active sensors can measure depth where the lack of texture makes event cameras uninformative, we took inspiration from our previous work Active Stereo Without Pattern Projector (ICCV 2023) to integrating a stereo event camera with an active sensor -- e.g., a LiDAR.

Inserting fake events inside the event stacks of the event-stereo network (**VSH**) or even directly inside the event stereo stream (**BTH**), we managed to alleviate the aforementioned issues.



Overview of a generic event-based stereo network and our hallucination strategies. State-of-the-art event-stereo frameworks (a) pre-process raw events to obtain event stacks fed to a deep network. In case the stacks are accessible, we define the model as a gray box, otherwise as a black box. In the former case (b), we can hallucinate patterns directly on it (VSH). When dealing with a black box (c), we can hallucinate raw events that will be processed to obtain the stacks (BTH).

Contributions:

- We prove that LiDAR-stereo fusion frameworks can effectively be adapted to the event stereo domain.
- Our VSH and BTH frameworks are general and work effectively with any structured representation among the eight we surveyed.
- Our strategies outperform existing alternatives inherited from RGB stereo literature on DSEC and M3ED datasets.
- VSH and BTH can exploit even outdated LiDAR data to increase the event stream distinctiveness and ease matching, preserving the microsecond resolution of event cameras and eliminating the need for synchronous processing dictated by the constant framerate of the depth sensor.

:fountain_pen: If you find this code useful in your research, please cite:

```
@inproceedings{bartolomei2024lidar,
  title={LiDAR-Event Stereo Fusion with Hallucinations},
  author={Bartolomei, Luca and Poggi, Matteo and Conti, Andrea and
Mattoccia, Stefano},
  booktitle={European Conference on Computer Vision (ECCV)},
  year={2024},
}
```

Pretrained Models

Here, you can download the weights of the baseline architecture trained on DSEC with eight different stacking representations.

To use these weights, please follow these steps:

1. Install GDown python package: `pip install gdown`
2. Download all weights from our drive : `gdown --folder <folder_link>` (`<folder_link>` will be provided soon)

Code

The **Test** section provides scripts to evaluate disparity estimation models on **DSEC** and **M3ED** datasets. It helps assess the accuracy of the models and saves predicted disparity maps.

Please refer to each section for detailed instructions on setup and execution.

Warning:

- Please be aware that we will not be releasing the training code for deep stereo models. The provided code focuses on evaluation and demonstration purposes only.
- With the latest updates in PyTorch, slight variations in the quantitative results compared to the numbers reported in the paper may occur.

:hammer_and_wrench: Setup Instructions

1. **Dependencies**: Ensure that you have installed all the necessary dependencies. The list of dependencies can be found in the `./requirements.txt` file.
2. **Build deform_conv**:

- Activate your virtual env
- `cd ./src/components/models/baseline/deform_conv/`
- `setup.py build_ext --inplace`

Datasets

We used seven datasets for training and evaluation.

DSEC

Download DSEC (`train_events.zip` `train_disparity.zip` and `train_calibration.zip`)

Next, we will provide soon preprocessed DSEC raw LiDAR scans:

```
$ cd PATH_TO_DSEC
$ gdown <folder_link>
$ unzip dsec_raw.zip
```

After that, you will get a data structure as follows:

```
dsec
├── train
│   ├── interlaken_00_c
│   │   ├── calibration
```

```
│   │   │   ├── cam_to_cam.yaml
│   │   │   └── cam_to_lidar.yaml
│   │   ├── disparity
│   │   │   ├── event
│   │   │   │   ├── 000000.png
│   │   │   │   ├── ...
│   │   │   │   └── 000536.png
│   │   │   ├── raw
│   │   │   │   ├── 000000.png
│   │   │   │   ├── ...
│   │   │   │   └── 000268.png
│   │   │   ├── raw_mae.txt
│   │   │   ├── raw_mae.png
│   │   │   ├── raw_bad1.txt
│   │   │   └── timestamps.txt
│   │   └── events
│   │       ├── left
│   │       │   ├── events.h5
│   │       │   └── rectify_map.h5
│   │       └── right
│   │           ├── events.h5
│   │           └── rectify_map.h5
│   ├── ...
│   └── zurich_city_11_c                # same structure as
train/interlaken_00_c
└── test
    └── ...
```

We managed to extract the raw LiDAR scans using only data from the official website. We used FasterLIO to de-skew raw LiDAR scans and Open3D to perform ICP registration.

M3ED

We will provide soon the preprocessing script for M3ED.

We managed to extract raw LiDAR using only data from the official website.

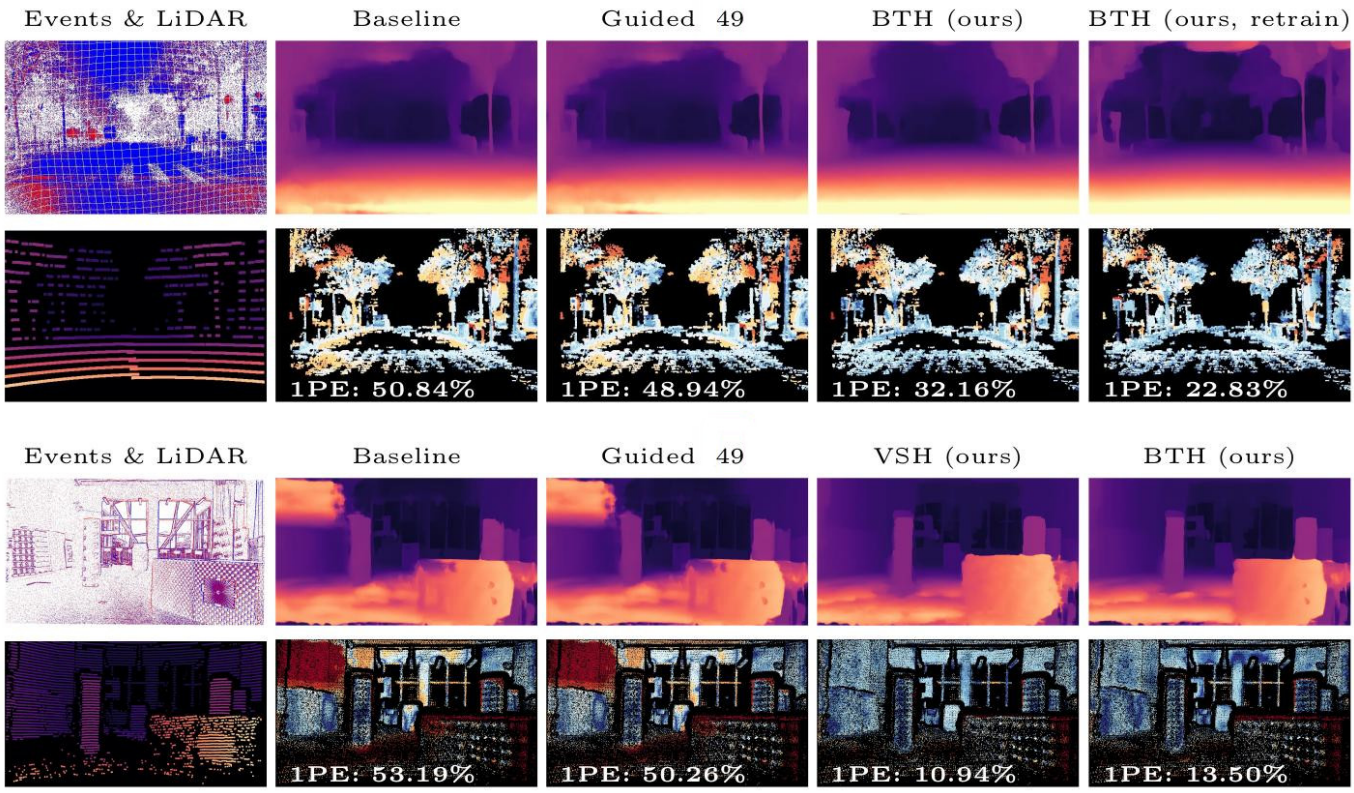
🚀 Test

This code snippet allows you to evaluate the disparity maps on DSEC and M3ED datasets. By executing the provided script, you can assess the accuracy of disparity estimation models on these datasets.

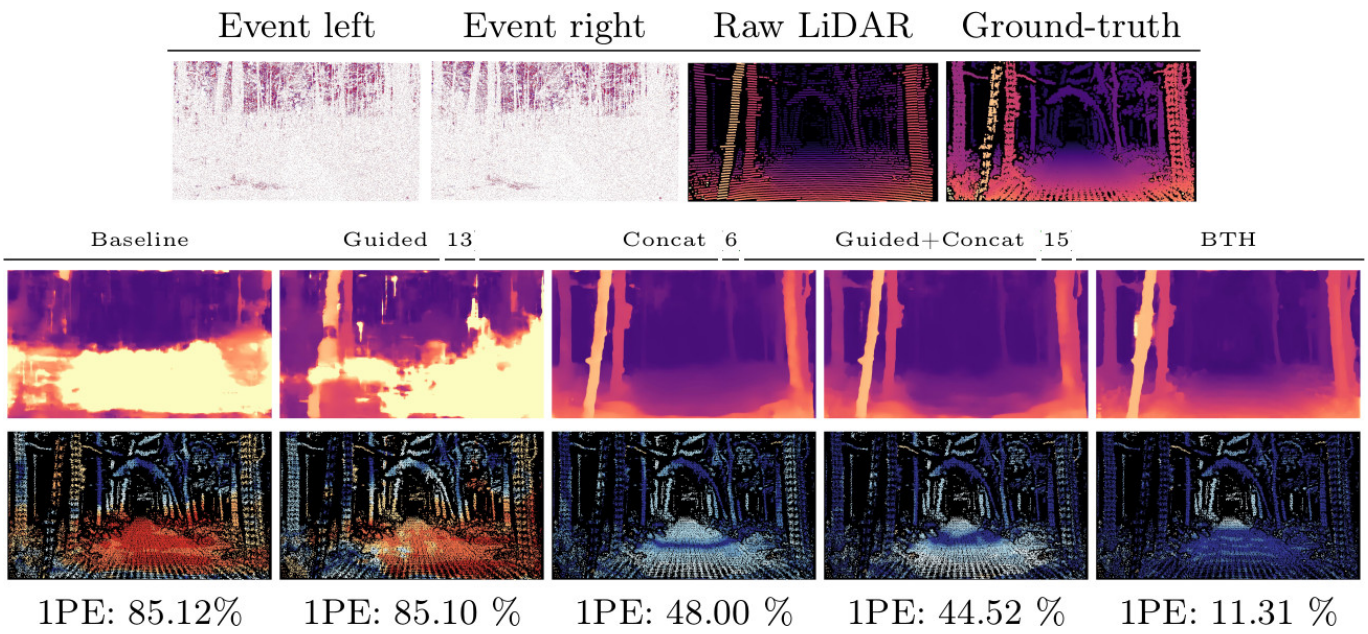
We will provide soon the code for evaluation.

🎨 Qualitative Results

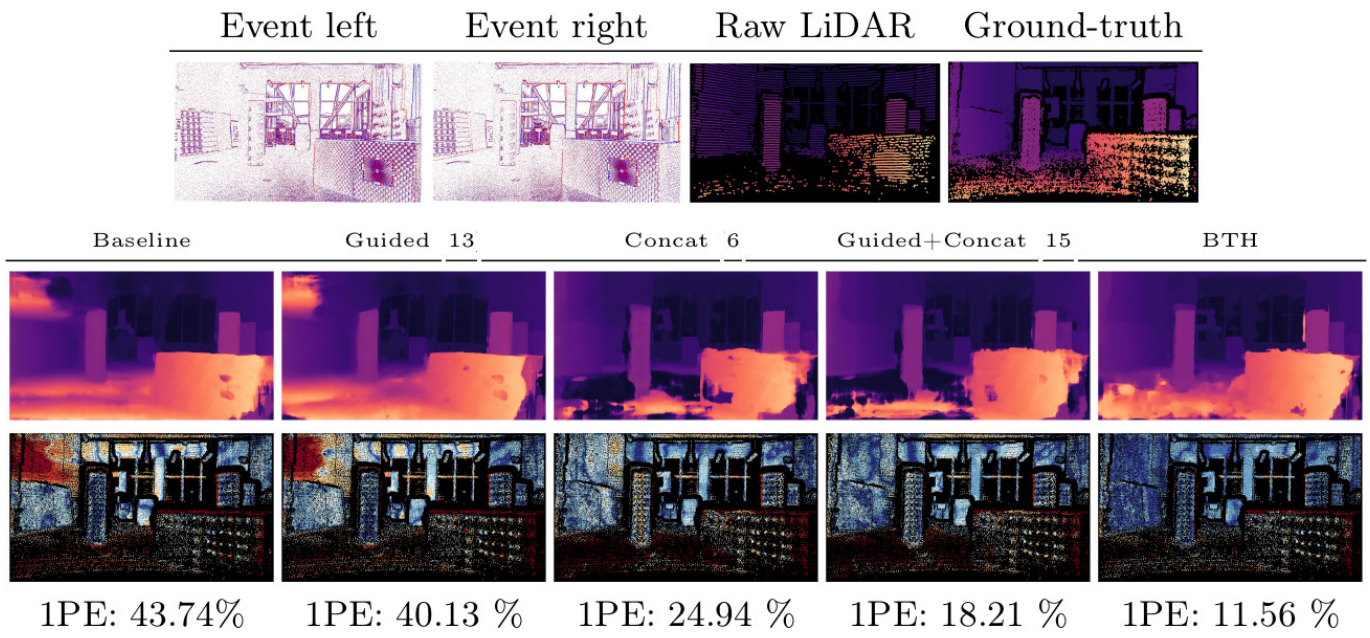
In this section, we present illustrative examples that demonstrate the effectiveness of our proposal.



Performance against competitors -- pre-trained models. On DSEC (top), BTH dramatically improves results over the baseline and Guided, yet cannot fully recover some details in the scene except when retraining the stereo backbone. On M3ED (bottom), both VSH and BTH with pre-trained models reduce the error by 5×.



Performance against competitors -- refined models (outdoor). Concat and Guided+Concat can reduce the error by about 40%, yet far behind the improvement yielded by BTH (more than 70% error rate reduction).



Performance against competitors -- refined models (indoor). our proposal confirms again the best solution for exploiting raw LiDAR measurements and improve the accuracy of event-based stereo networks.

✉️ Contacts

For questions, please send an email to luca.bartolomei5@unibo.it

🙏 Acknowledgements

We would like to extend our sincere appreciation to the authors of the following projects for making their code available, which we have utilized in our work:

- We would like to thank the authors of SE-CFF for providing their code, which has been instrumental in our stereo matching experiments.