# DAOVI: Distortion-Aware Omnidirectional Video Inpainting


†Ryosuke Seshimo, †Mariko Isogawa †Keio University

Contact: seshimo.ryosuke@keio.jp (Seshimo), mariko.isogawa@keio.jp (Isogawa)


Project Page

## Research Overview

### Our Goal

Remove regions specified by the user from **omnidirectional video** and complete the content seamlessly using **video inpainting**.

Distortions in Omnidirectional Video

### Background

- Omnidirectional video often includes unintended objects because of the wide field of view. Video inpainting is effective for removing such objects.
- Distortion varies with spatial coordinates in omnidirectional video.
- Because existing video inpainting methods are designed for ordinary videos with a narrow field of view, they are not well suited to omnidirectional video with spatially varying distortion.

### Contributions

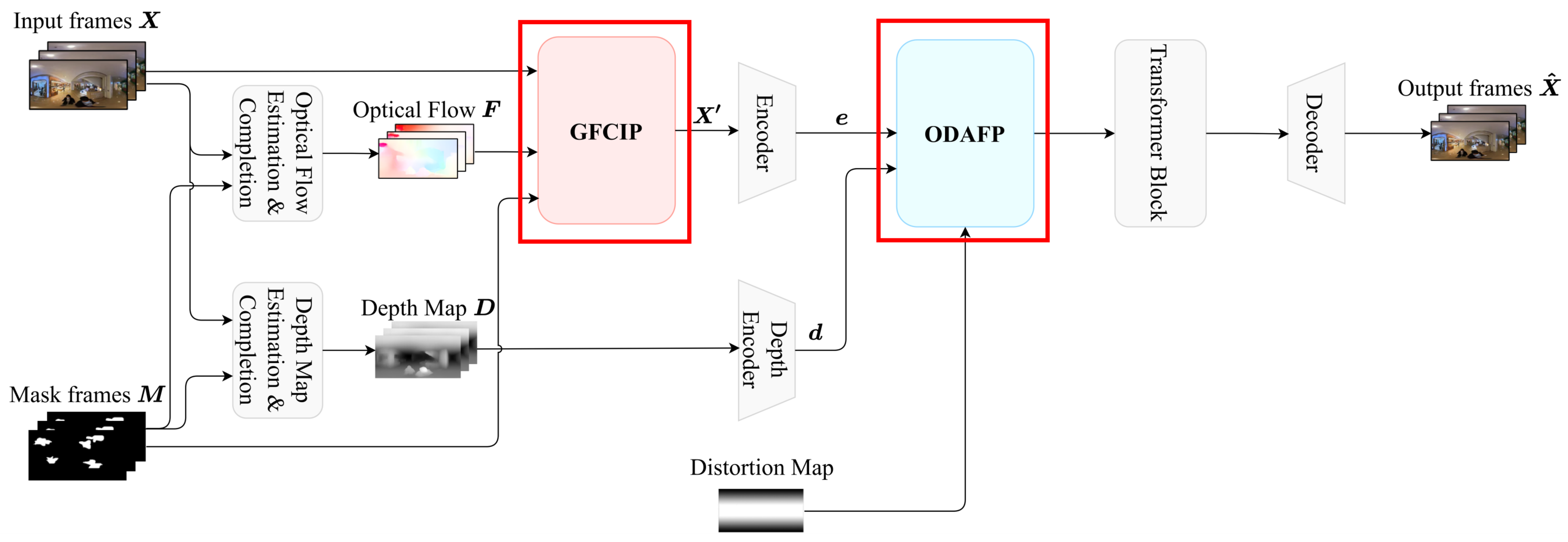
- We propose a **deep learning-based video inpainting method for omnidirectional videos**.
- Introducing a module based on geodesic distance improves performance without adding learnable parameters. (**GFCIP module**)
- Distortion-aware inpainting using a Distortion Map and distortion-adaptive convolutions. (**ODAFP module**)

## Proposed Method

Input : Video frames $\boldsymbol{X}$,  Mask frames $\boldsymbol{M}$

Output : Inpainted video frames $\hat{\boldsymbol{X}}$

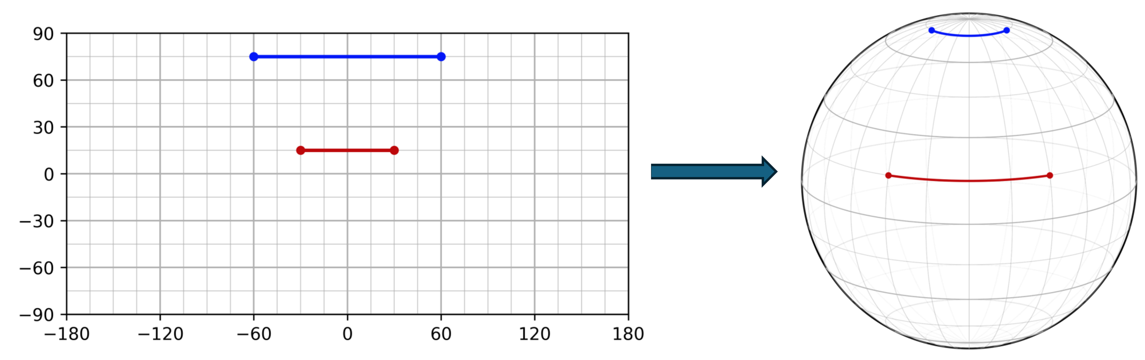
1. **Estimate optical flow** $\boldsymbol{F}$ from video frames $\boldsymbol{X}$.
2. **In image space**, the **GFCIP** module partially inpaints the masked region using optical flow $\boldsymbol{F}$.
3. **In feature space**, the **ODAFP** module inpaints the remaining masked region.
4. The decoder outputs inpainted video frames $\hat{\boldsymbol{X}}$.




## Proposed Modules

### Geodesic Flow-Consistent Image Propagation (GFCIP)

1. For each pixel in the masked region, compute the forward and backward optical flow consistency error.
2. Convert this error to **geodesic distance** on the sphere.
3. Inpaint only the parts whose error is below a threshold, using pixel values from the adjacent frame.



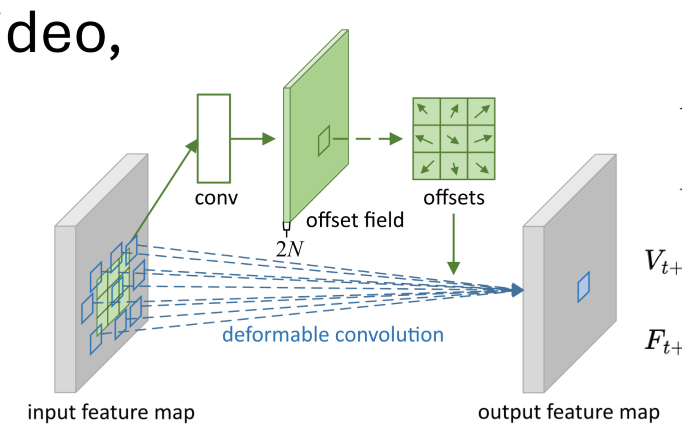
Convert to geodesic distance

### Omnidirectional Depth-Assisted Feature Propagation (ODAFP)

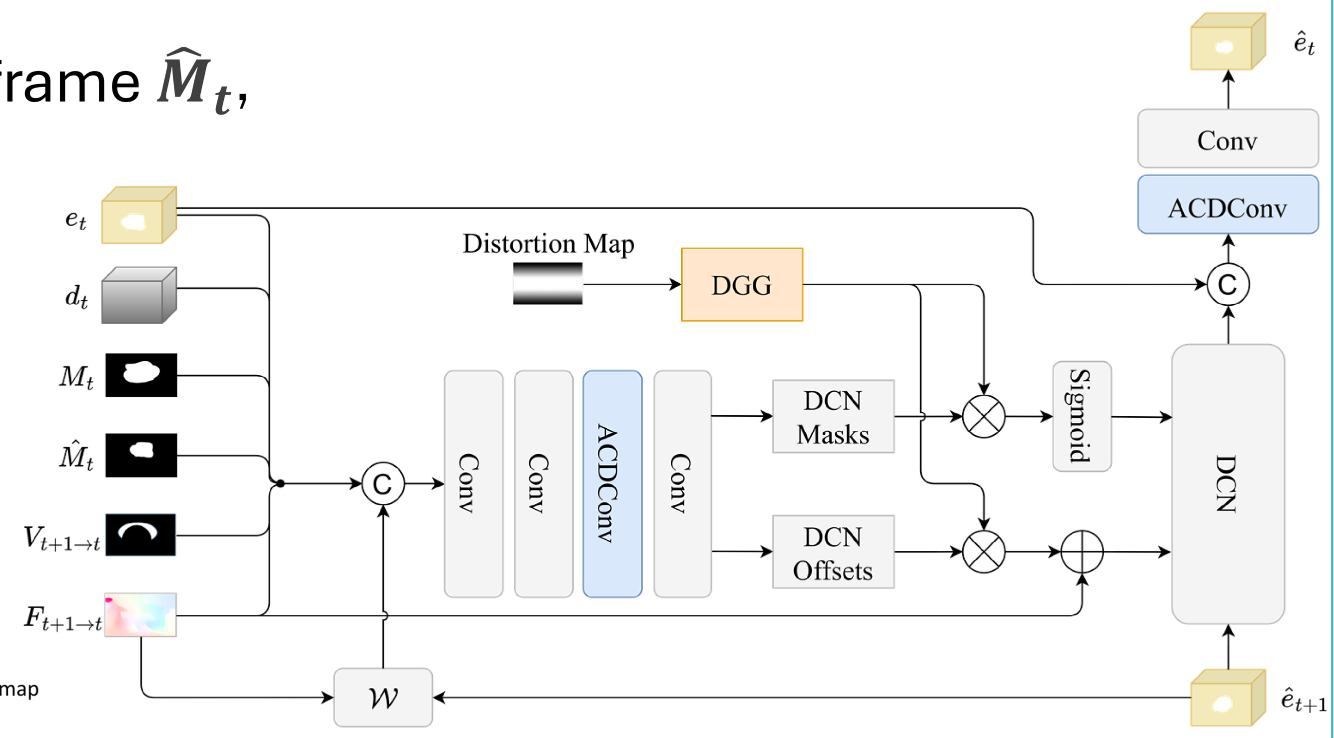
Input : Feature map $\boldsymbol{e}_t$, Depth feature map $\boldsymbol{d}_t$, Mask frame $\boldsymbol{M}_t$, Unfiled region mask frame $\widehat{\boldsymbol{M}}_t$, Optical flow $\boldsymbol{F}_{t+1\to t}$, Validity map $\boldsymbol{V}_{t+1\to t}$, Adjacent output of ODAFP $\hat{\boldsymbol{e}}_{t+1}$

Output : Inpainted feature maps $\hat{\boldsymbol{e}}_t$

1. Using **ACDConv**[1], which is designed for omnidirectional video, compute the **DCN[2] masks and offsets**.
2. **Modulate the masks and offsets** with the **Distortion Map**[3] encoding omnidirectional distortion.
3. Apply DCN to adjacent output of ODAFP $\hat{\boldsymbol{e}}_{t+1}$.




DCN




## Experiment

### Dataset : ODV360[4]

- Real world dataset covering indoor and outdoor scenes.
- We generate and use randomly shaped mask frames.

| Method | PSNR[dB](↑) | SSIM(↑) | WS-PSNR[dB](↑) | WS-SSIM(↑) | VFID(↓) |
|---|---|---|---|---|---|
| FuseFormer [5] | 29.18 | 0.9727 | 28.87 | 0.9392 | 0.320 |
| STTN [6] | 32.21 | 0.9815 | 31.09 | 0.9520 | 0.238 |
| ProPainter [7] | 32.81 | 0.9877 | 32.86 | 0.9648 | 0.149 |
| Ours | **33.37** | **0.9888** | **33.37** | **0.9661** | **0.138** |

### Quantitative Results

Outperforms existing methods on every metric.

### Qualitative Results

Produces inpainted results with fewer artifacts than existing method.

GT  ProPainter  Ours

## References


[1] Chuanqing Zhuang *et al.*, "ACDNet: Adaptively Combined Dilated Convolution for Monocular Panorama Depth Estimation", AAAI, 2022.
[2] Jifeng Dai *et al.*, "Deformable Convolutional Networks", ICCV, 2017.
[3] Yule Sun *et al.*, "Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video", IEEE Signal Processing Letters, 2017.
[4] Mingdeng Cao *et al.*, "Ntire 2023 challenge on 360° omnidirectional image and videosuper-resolution: Datasets, methods and results", CVPRW, 2023.
[5] Rui Liu *et al.*, "Fuseformer: Fusing fine-grained information intransformers for video inpainting", ICCV, 2021.
[6] Yanhong Zeng *et al.*, "Learning joint spatial-temporal transformations for video inpainting", ECCV, 2020.
[7] Shangchen Zhou *et al.*, "ProPainter: Improving propagation and transformer for video inpainting", ICCV, 2023.