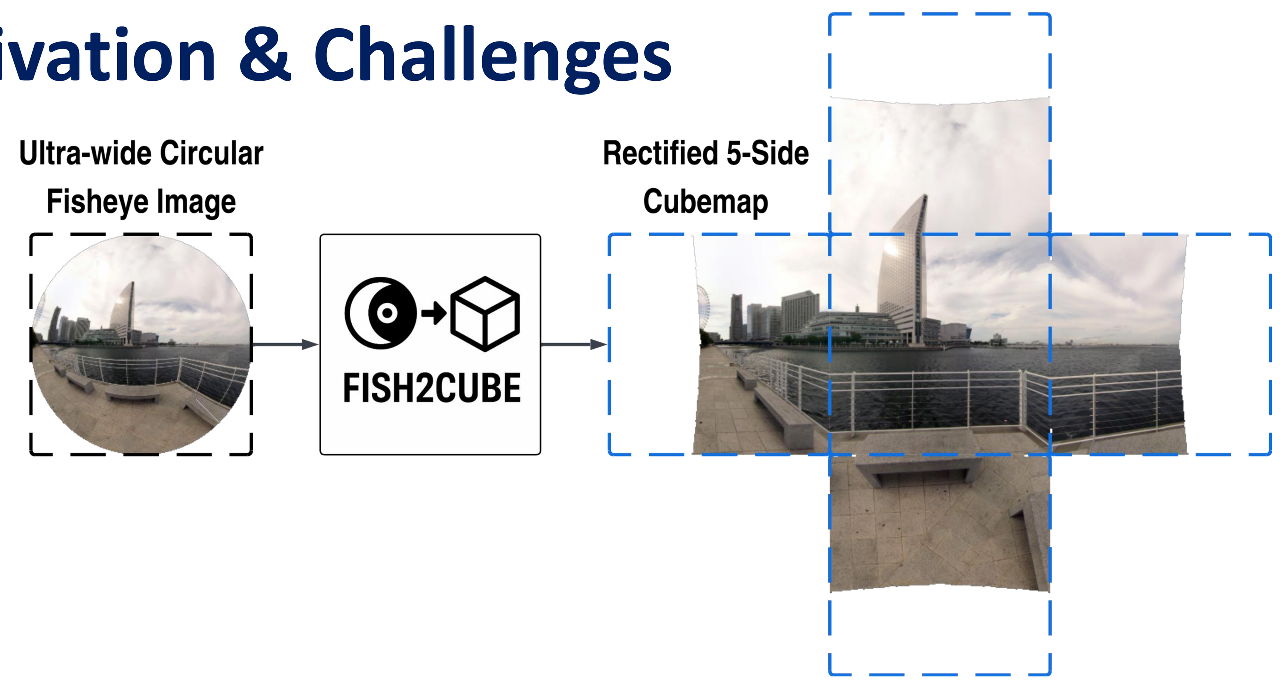
Split-and-Combine Rectification of Ultra-Wide Fisheye Images into Cubemaps

Yuang Fan, Xuhai Xu, and Xiaofan Jiang
Columbia University
{yf2676, xx2489}@columbia.edu, jiang@ee.columbia.edu

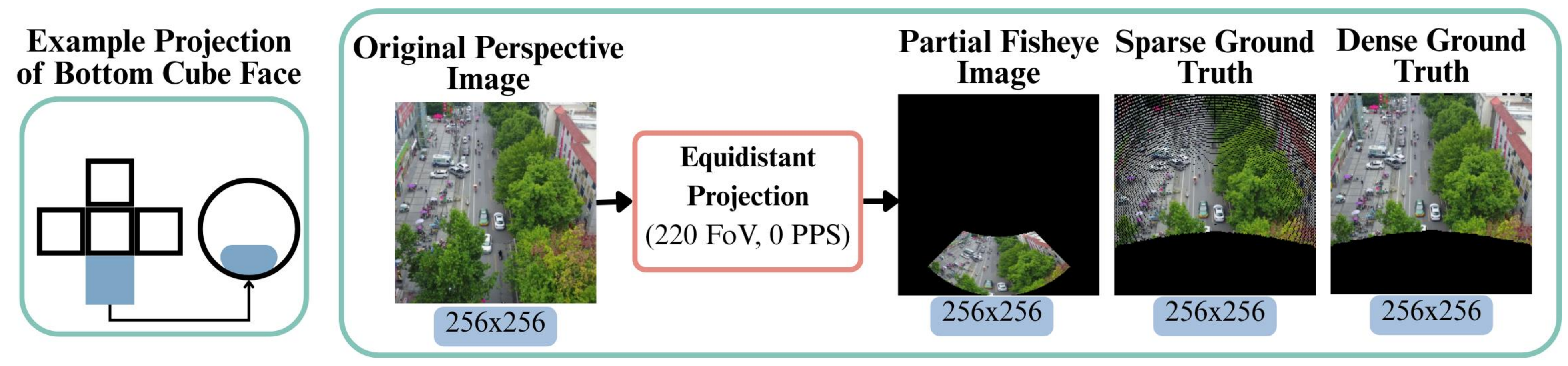
I. Motivation & Challenges



Ultra-wide fisheye cameras **(FoV > 180°)** capture panoramic scenes for drones, vehicles, and surveillance — but suffer from severe geometric distortion.
Challenge: Existing geometric and learning-based rectification methods are limited to moderate FoVs (90°–120°) and fail on ultra-wide fisheye images (>180°), where single-perspective mappings inherently cause severe distortion or information loss.
We address this challenge by proposing a ***split-and-combine*** framework that rectifies fisheye images into **five cubemap faces**, each processed and enhanced independently.

II. Data Preparation

Challenge: Supervised training is hindered by the scarcity of real fisheye–cubemap pairs and the difficulty of simulating >180° distortions, which require multi-view synthesis and lead to under sampled, low-detail rectifications.



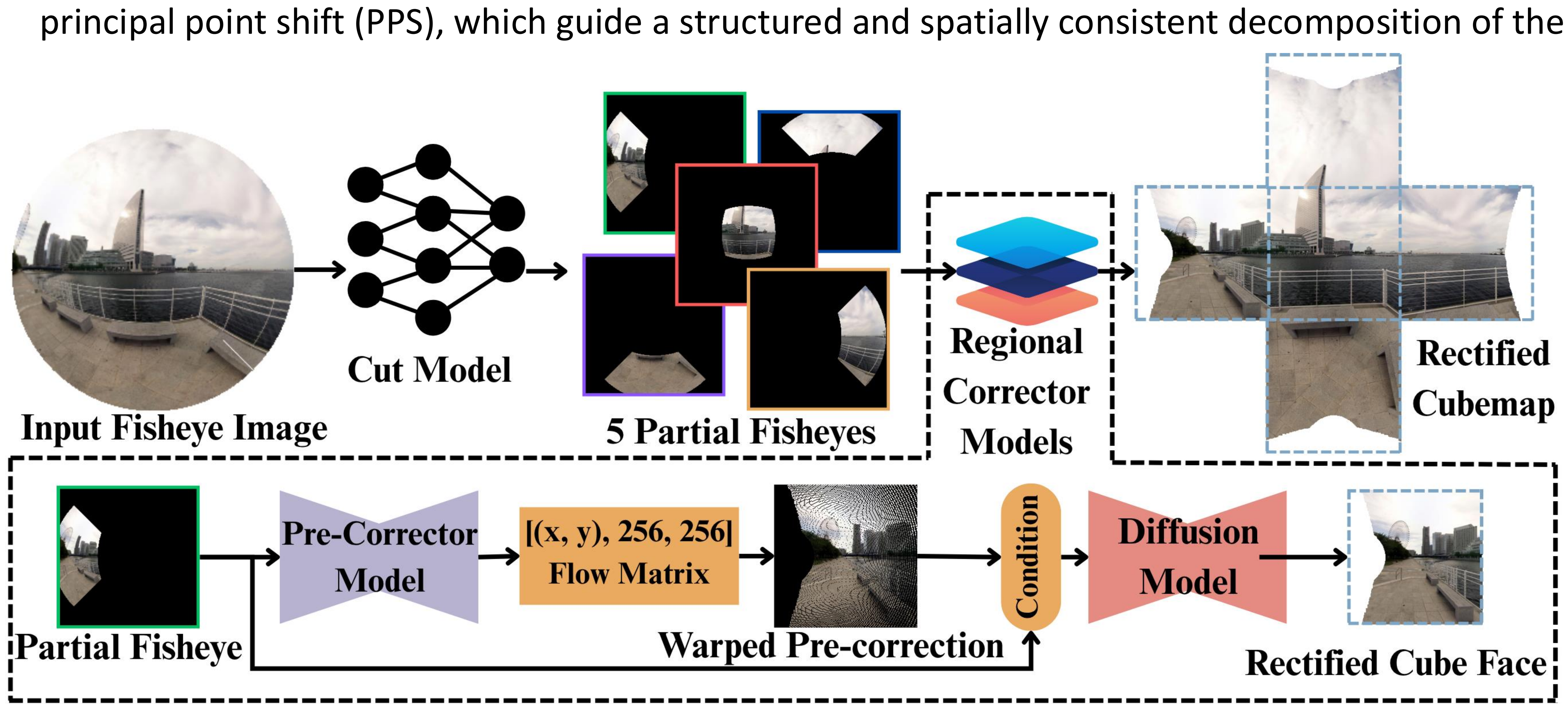
We synthesize **partial-fisheye** training data from **VisDrone** images:

- Project perspective images into synthetic fisheye using equidistant model (FoV 180°–240°., 0-3% Principle Point Shift).
- Generate sparse & dense supervision targets for each cube face.

This enables **partial supervision** for each regional corrector.

II. Split & Combine Framework

- Our system first divides each ultra-wide fisheye image into five directional regions—front, left, right, top, and bottom—and processes these regions independently through specialized rectifiers before seamlessly reassembling them into a coherent cubemap representation.
- A lightweight **Cut Model**, trained even with limited full fisheye-cubemap data, predicts key geometric parameters such as the field of view (FoV) and principal point shift (PPS), which guide a structured and spatially consistent decomposition of the fisheye image into its directional components.



- Each region is processed by a two-stage corrector:
  1. **Flow-based Pre-corrector**: A U-Net with attention predicts dense warping flow.
  2. **Diffusion-based Enhancer**: A U-Net based diffusion model refines missing details with hybrid loss (noise + L1 + VGG).
- Front and peripheral regions use separate rectifiers for tailored correction.
- Training proceeds in three stages:

Pretrain Flow Network → **Train Diffusion Model** → **Joint Fine-tuning**

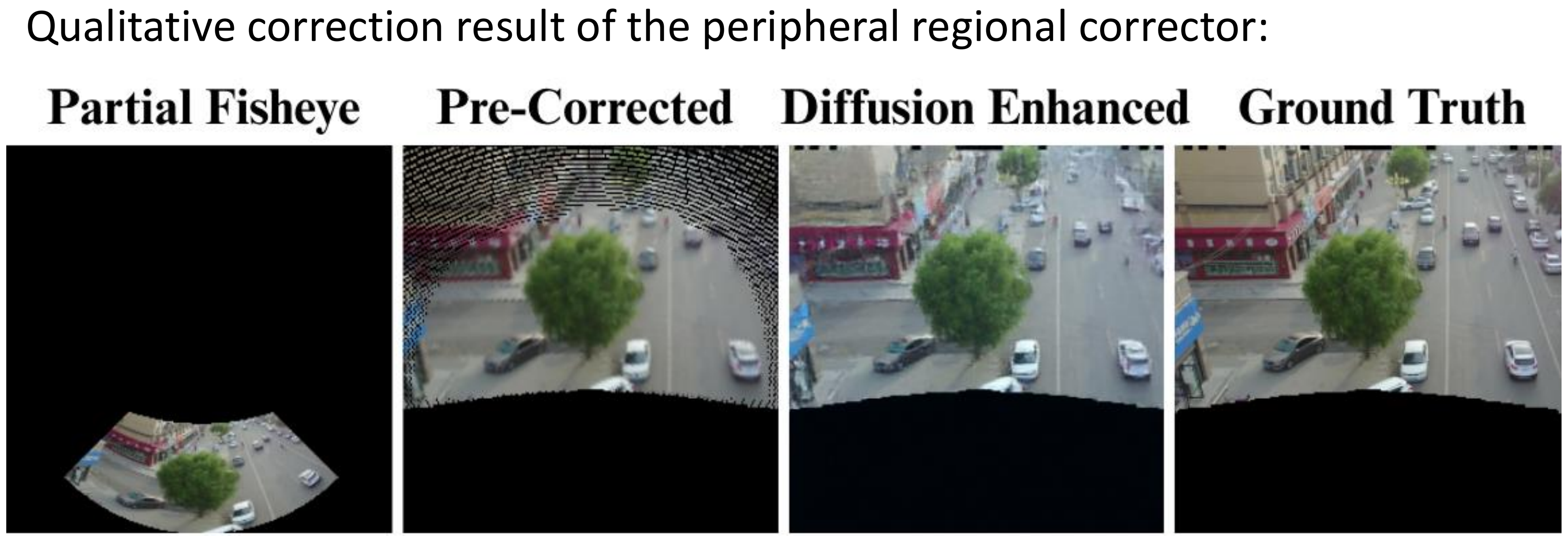
III. Results

We evaluate on synthetic VisDrone validation data (FoV > 180°).

Despite the harder setting, our regional correctors achieve competitive fidelity:

Method	PSNR ↑	SSIM ↑
PCN	24.98	0.87
SimFIR	22.47	0.86
DDA	24.60	0.76
Ours	22.53	0.82

Qualitative correction result of the peripheral regional corrector:



IV. Conclusion & Future Work

Summary: Modular split-and-combine rectification resolves geometric limits of single-perspective projection.
Strengths: interpretable parameters, partial supervision, strong visual fidelity.

🔜 **Next:**
❑ Cross-face consistency refinement
❑ Support for wider lens models
❑ Real-time edge deployment

Acknowledgment

This work was supported by the National Science Foundation (NSF) and Center for Smart Streetscapes (CS3) under NSF Cooperative Agreement No. EEC-2133516. The views expressed are those of the authors and do not necessarily reflect the policies or endorsements of Columbia University, NSF, or the U.S. Government.