Deep-BrownConrady: Prediction of Camera Calibration and Distortion Parameters Using Deep Learning and Synthetic Data

AILiveSim

Faiz Muhammad Chaudhry

Jarno Ralli (AILiveSim), Jerome Leudet (AILiveSim), Fahad Sohrab (Tampere University), Farhad Pakdaman (Tampere University), Pierre Corbani (AILiveSim), Moncef Gabbouj (Tampere University)

IEEE Transactions on Automation Science and Engineering, 2025

Why camera calibration matters

- Core for autonomous driving, robotics, AR, industrial systems
- Needed for
    - 3D reconstruction
    - Object detection and tracking
    - Accurate overlays and navigation
- But: large vision datasets usually have no calibration data

Fig. Distortion effects on a city scene from AILiveSim with 90° H-FOV with k1 = 0.25

Limitations of Traditional Calibration

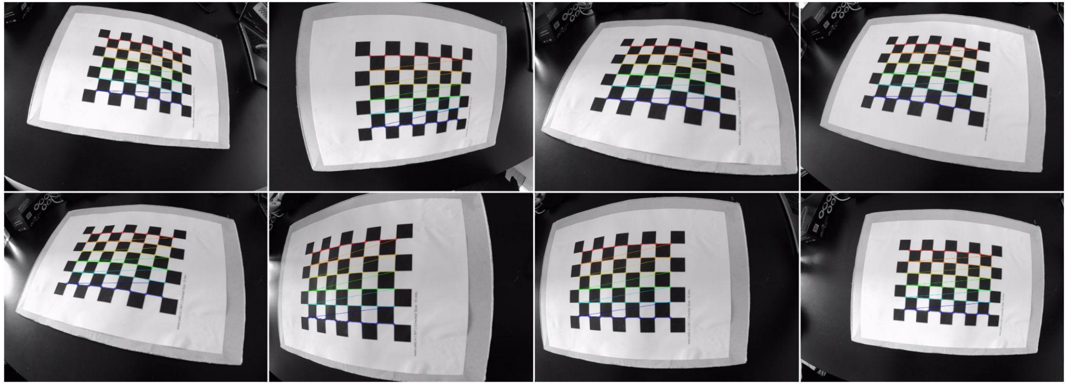


Fig. Samples of photos taken by an uncalibrated camera

- Needs multiple images of checkerboards or dot patterns
- Requires controlled capture process
- Hard to redo frequently in the field
- Not applicable to existing image only datasets

Goal and Contributions

- Predict camera parameters from a single RGB image
- Parameters
  - H-FOV
  - Principal point (cx, cy)
  - Brown-Conrady distortion (k1, k2, k3, p1, p2)
- Main contributions
  - Large synthetic dataset generated using AILiveSim
  - ResNet50 based Deep-BrownConrady model
  - Strong results on KITTI and other real datasets

Fig. Image sample from KITTI test dataset

Camera and Distortion Model

- Pinhole model
  - Intrinsics K: fx, fy, cx, cy
  - Projection from 3D to 2D
- Brown-Conrady distortion
  - Radial: k1, k2, k3
  - Tangential: p1, p2
- We predict H-FOV instead of fx, fy directly

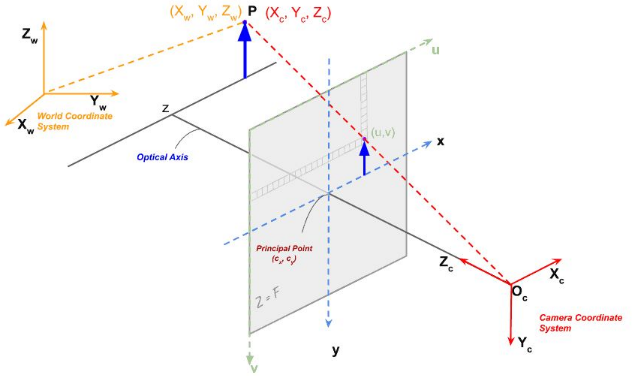
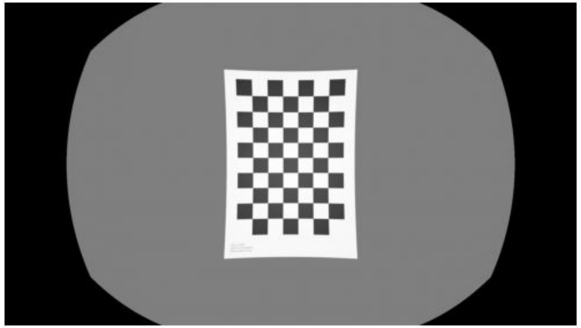
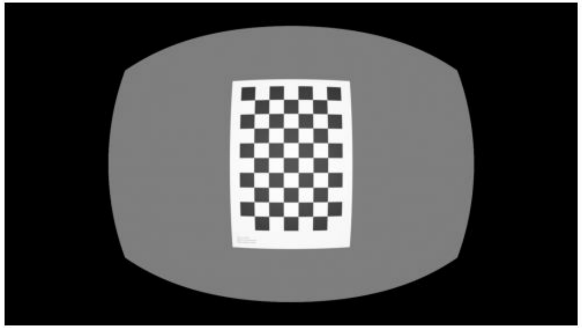


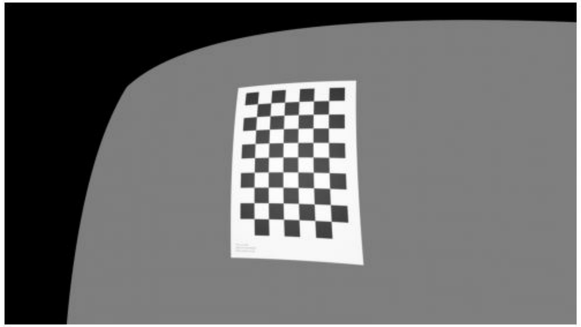
Fig. Projection of point P onto the image plane using the Pinhole Camera Model [15].



(a) Barrel distortion



(b) Pincushion distortion



(c) Tangential distortion

Fig. Lens distortions: (a) Barrel distortion with k1 = −0.5, (b) Pincushion distortion with k1 = 0.5, and (c) Tangential distortion with p1 and p2 set to 0.1

AILiveSim and Camera Parameter Search (CPS) Dataset

- AILiveSim: Unreal based simulator with full control
- CPS dataset
  - 1.495 million synthetic images
  - FOV from 30 to 150 degrees
  - Resolutions: 1920×1080 and 1392×512
- Rich urban scenes with many straight lines

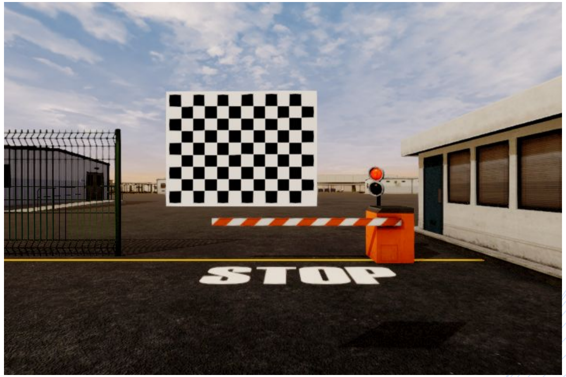


Fig. Camera Calibration inside AILiveSim

Fig. Image generated using AILiveSim with 150° H-FOV and no distortion

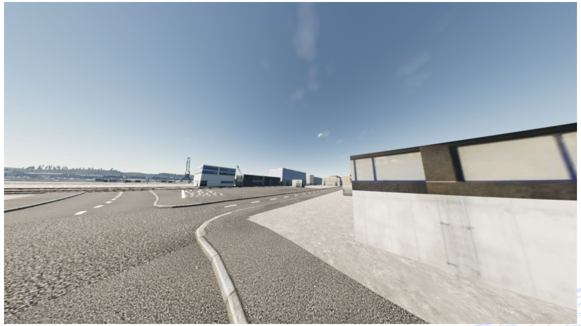
Dataset Summary

- Synthetic Camera Parameter Search Data
  - 1.495M images, multiple FOV (30-150) and resolutions (1920x1080, 1392x512)
- Real data
  - 10k images from KITTI for training
  - Additional KITTI, Malaga, Cityscapes for testing
- Goal
  - Learn from synthetic + real
  - Validate on real

Dataset Summary

a) H - FOV = 29.04° , k_1 = 0.384, k_2 = 0.465, k_3 = 0.076, p_1 = 0.0167, p_2 = 0.0326, c_x = 947, c_y = 523

b) H - FOV = 88.18° , k_1 = 0.0027, k_2 = - 0.0010, k_3 = 0.0010, p_1 = -0.0053, p_2 = 0.0002, c_x = 981, c_y = 538



c) H - FOV = 147.57° , k_1 = 0.0006, k_2 = - 0.0001, k_3 = - 0.000002, p_1 = - 0.0025, p_2 = - 0.0030, c_x = 948, c_y = 545

Fig. Sample distorted images from the CPS dataset with varying horizontal fields of view (H-FOV). Each image includes Brown-Conrady lens distortion parameters (k1, k2, k3, p1, p2) and principal point coordinates (cx, cy), demonstrating the range and diversity of calibration settings used during dataset generation

DBC Model Architecture

- Backbone
  - ResNet50 feature extractor
- Regression head
  - Fully connected layers
  - Outputs 8 parameters
- Loss
  - Mean Squared Error on all parameters

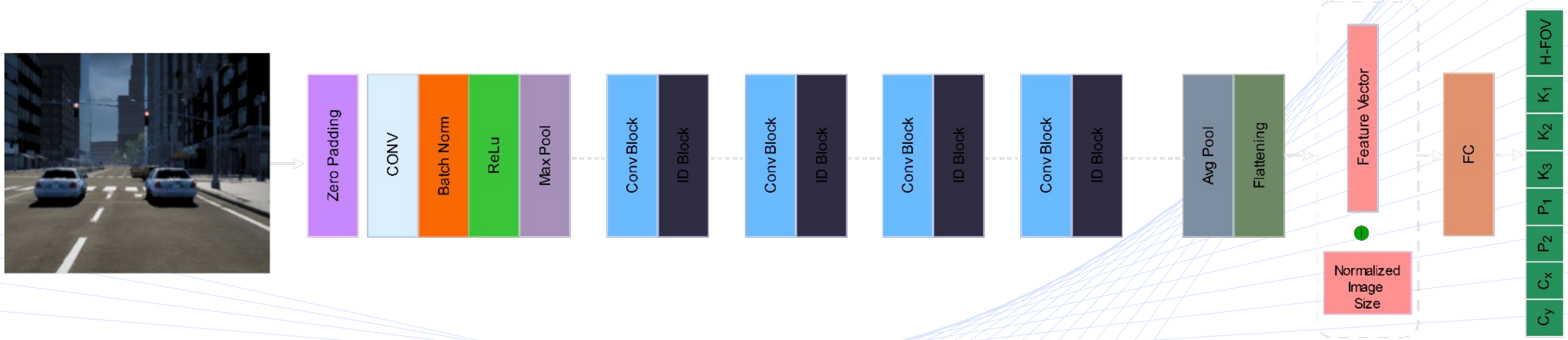


Fig. Architecture of the DBC v3 model, based on the ResNet50 framework

From DBC v1 to DBC v3

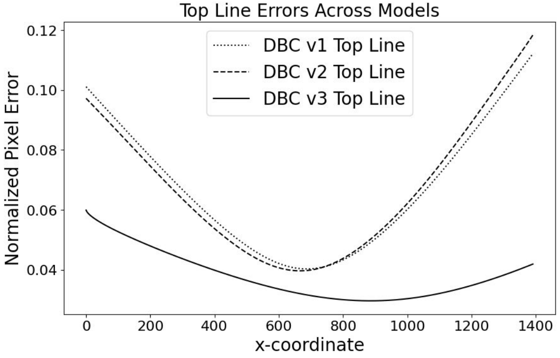
- DBC v1
  - Single resolution, basic ResNet50 regression
- DBC v2
  - Supports two resolutions
  - Grouped batches per resolution
- DBC v3
  - Adds normalized image width and height as extra features
  - Best performance and generalization

Training Setup

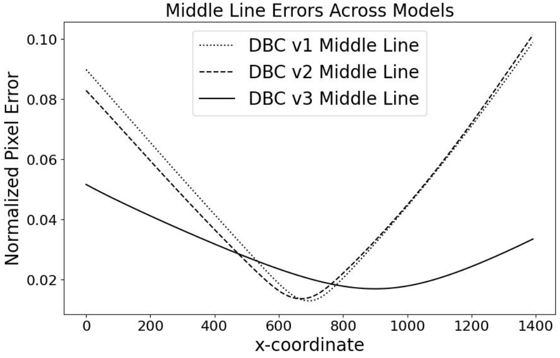
- Loss
  - MSE on 8 parameters
- Optimizer
  - AdamW, learning rate schedule with decay
- Batch size 128, 60 epochs
- Data splits
  - 70 percent train, 15 percent validation, 15 percent test
- Robustness tests
  - Gaussian blur, motion blur, gamma, pixel dropout

Results: Spatial Error Maps

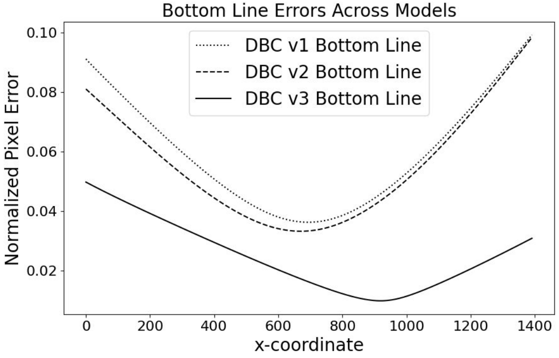
- For each pixel: compare position after ground truth vs predicted undistortion
- Compute normalized Euclidean error
- Average over 5000 test images
- Extract top, middle, bottom lines for analysis



(a) Top Horizontal Line Errors



(b) Center Horizontal Line Errors



(c) Bottom Horizontal Line Errors

Fig. Normalized pixel-wise errors along top, middle, and bottom horizontal lines for DBC v1, DBC v2, and DBC v3. Each graph shows normalized pixel-wise errors across x-coordinates for all three models.

Results: Line Grids

- Synthetic line grid experiment
- Distort with ground truth parameters
- Undistort with predicted parameters

(a) Original

(b) Distorted

(c) DBC v1

(d) DBC v2

(e) DBC v3

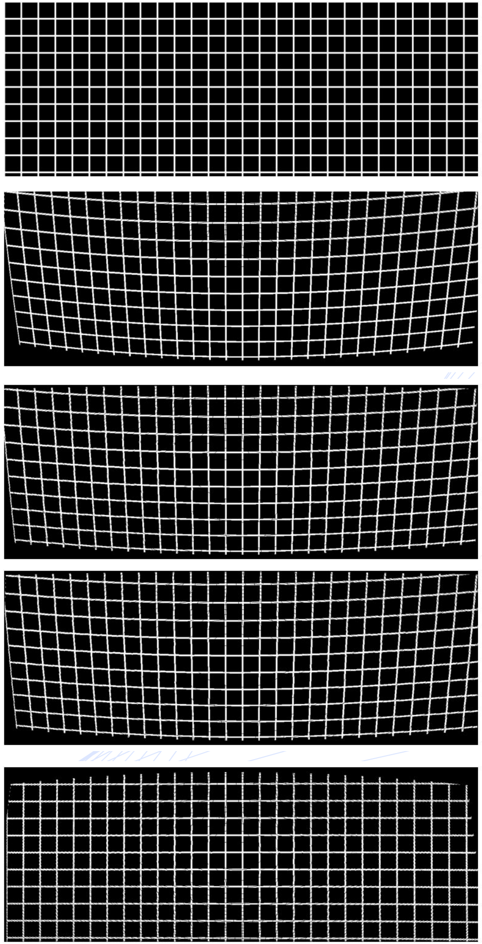


Fig. Visualization of the distortion and undistortion process. (a) Original line image, (b) Image distorted using true distortion parameters, (c) Undistorted image using predicted parameters from DBC v1, (d) Undistorted image using DBC v2, (e) Undistorted image using DBC v3.

Results: Traditional Calibration vs DBC

- Checkerboard calibration using few images
- Our model uses one arbitrary scene image
- With 1 to 3 checkerboard images
  - Traditional method can be less accurate
  - Sensitive to pattern placement

TABLE II

COMPARISON OF THE TRADITIONAL CAMERA CALIBRATION APPROACH WITH DBC V1, DBC V2, AND DBC V3: THE TRADITIONAL CALIBRATION METHOD RELIED ON IMAGES OF A KNOWN CALIBRATION OBJECT, WHEREAS THE DEEP-BROWNCONRADY MODELS UTILIZED A SINGLE IMAGE CAPTURED IN THE WILD, WITHOUT ANY OBJECTS OF KNOWN SHAPE OR SIZE

Method	TOP HORIZONTAL LINE		MIDDLE HORIZONTAL LINE		BOTTOM HORIZONTAL LINE	
	MIN ERROR	MAX ERROR	MIN ERROR	MAX ERROR	MIN ERROR	MAX ERROR
DBC v1	34.96e-03	129.66e-03	9.01e-03	108.23e-03	9.01e-03	116.89e-03
DBC v2	13.47e-03	70.92e-03	16.68e-03	56.18e-03	16.68e-03	62.98e-03
DBC v3	5.84e-03	41.20e-03	23.65e-03	44.12e-03	23.65e-03	49.73e-03
Calibration, 1 image	186.13e-03	753.72e-03	122.22e-03	767.03e-03	203.73e-03	776.89e-03
Calibration, 3 images	0.09e-03	33.22e-03	0.39e-03	26.08e-03	0.18e-03	37.48e-03
Calibration, 5 images	2.97e-03	27.36e-03	3.31e-03	22.93e-03	2.93e-03	30.25e-03
Calibration, 7 images	5.60e-03	12.42e-03	5.42e-03	10.96e-03	5.57e-03	13.77e-03
Calibration, 9 images	2.69e-03	9.31e-03	3.69e-03	7.70e-03	2.05e-03	10.74e-03

Results: AnyCalib Comparison

- AnyCalib
  - Single view calibration for intrinsics and radial distortion
  - Does not predict tangential distortion
- DBC v3
  - Lower MAE
  - Predicts all 5 distortion parameters
  - 27x faster inference

MAE COMPARISON BETWEEN DBC V3 AND ANYCALIB ACROSS THREE DATASETS. MISSING VALUES (-) INDICATE PARAMETERS NOT PREDICTED BY ANYCALIB. ALL VALUES ARE UNIT-NORMALIZED

Dataset	norm_cx	norm_cy	hfov	k1	k2	k3	p1	p2
DBC v3 (Ours)								
KITTI	**0.002**	**0.001**	**0.009**	**0.008**	**0.007**	**0.004**	**0.003**	**0.003**
Malaga	0.017	0.051	0.187	0.150	**0.076**	0.054	**0.016**	**0.007**
CityScapes	0.037	0.039	0.508	0.182	0.093	**0.053**	**0.031**	**0.004**
AnyCalib (2025)								
KITTI	0.005	0.045	0.153	0.228	0.571	1.053	–	–
Malaga	**0.009**	**0.023**	**0.102**	**0.070**	0.078	0.036	–	–
CityScapes	**0.028**	**0.004**	**0.200**	0.183	**0.837**	1.585	–	–

Limitations

- Limited public datasets with full calibration metadata
- Performance drops on novel resolutions outside training range
- Extreme FOV or aspect ratios remain challenging

Conclusion and Future Work

- Deep-BrownConrady
  - Predicts camera intrinsics and Brown-Conrady distortion from one image
- Outperforms
  - Traditional calibration with few images without objects of known geometry
  - AnyCalib in accuracy and speed
- Future work
  - Incorporate line or edge detectors
  - Explore graph based or hybrid architectures
  - Extend to more cameras and resolutions