Revisiting Calibration of Wide-Angle Radially Symmetric Cameras

Andrea Porfiri Dal Cin[1], Francesco Azzoni[1], Giacomo Boracchi[1], and Luca Magri[1]

DEIB, Politecnico di Milano, Italy
firstname.lastname@polimi.it

Abstract. Recent learning-based calibration methods yield promising results in estimating parameters for wide field-of-view cameras from single images. Yet, these end-to-end approaches are typically tethered to one fixed camera model, leading to issues: **(i)** lack of flexibility, necessitating network architectural changes and retraining when changing camera models; **(ii)** reduced accuracy, as a single model limits the diversity of cameras represented in the training data; **(iii)** restrictions in camera model selection, as learning-based methods need differentiable loss functions and, thus, undistortion equations with closed-form solutions. In response, we present a novel *two-step* calibration framework for radially symmetric cameras. Key to our approach is a specialized CNN that, given an input image, outputs an implicit camera representation (VaCR), mapping each image point to the direction of the 3D light ray projecting onto it. The VaCR is used in a subsequent robust non-linear optimization process to determine the camera parameters for any radially symmetric model provided as input. By disentangling the estimation of camera model parameters from the VaCR, which is based only on the assumption of radial symmetry in the model, we overcome the main limitations of end-to-end approaches. Experimental results demonstrate the advantages of the proposed framework compared to state-of-the-art methods. Code is at github.com/andreadalcin/RadiallySymmetricCalib.

1 Introduction

Camera calibration is crucial for various Computer Vision tasks, such as image rectification, 3D reconstruction, and visual localization. As wide field-of-view cameras become more prevalent, calibration techniques have increasingly targeted the intrinsic and distortion parameters of fisheye lenses, 360-degree cameras, and spherical cameras. This paper specifically addresses the calibration of radially symmetric cameras, encompassing these devices.

Calibration methods can be categorized into geometric-based [1, 3, 5, 11, 23, 27, 30, 32, 38, 39] and learning-based [4, 13, 14, 17, 18, 20, 24, 29, 34, 35]. Geometric-based approaches utilize checkerboard patterns or handcrafted features to estimate camera parameters, but often falter in unstructured settings or without

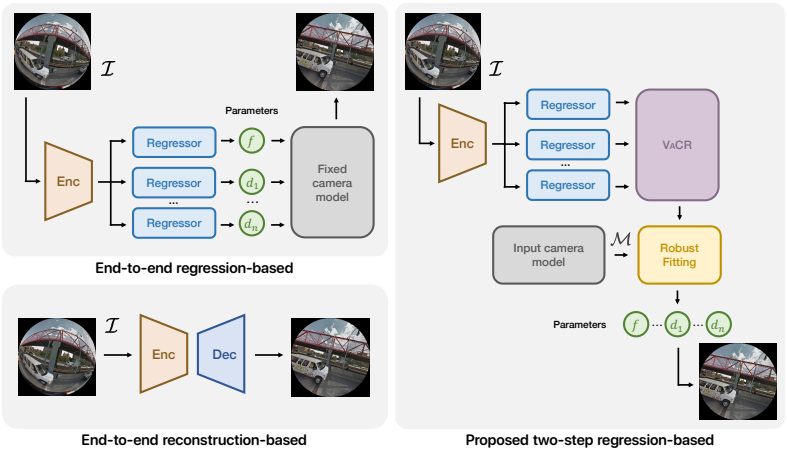

Fig. 1: Comparison with End-to-End Calibration Methods. Our two-step approach uses a CNN to regress the VaCR, the implicit camera representation driving a robust fitting procedure to estimate the parameters for the input camera model $\mathcal{M}$. This contrasts with end-to-end approaches, which either directly estimate parameters for a fixed camera model or produce undistorted images without intermediary steps.

manual intervention. Conversely, learning-based approaches can calibrate a camera without calibration objects and have shown promising results in estimating intrinsic and extrinsic parameters in uncontrolled settings.

Nonetheless, learning-based methods have practical limitations. **(i)** They typically regress camera parameters directly using an *end-to-end* approach that relies on a fixed camera model, which limits calibration accuracy when said model mismatches the types of cameras being calibrated. **(ii)** Transitioning to a more appropriate camera model requires architectural changes and retraining of the network with a dataset of distorted images based on the new camera model. **(iii)** Selecting an effective loss function for network training presents difficulties. The mainstream approach involves formulating the loss function based on the undistortion equations of the camera model, as seen in [24, 34]. However, differentiating the loss requires these equations to have a unique closed-form solution, which does not hold for all camera models. Thus, this limitation narrows the range of models suitable for this approach. Another strategy involves deriving the loss directly from camera parameters, as in [4], with metrics such as the root-mean-square error. Nevertheless, this approach is incompatible with camera models having inherent ambiguities, such as the Extended Unified (EUCM) [15] and Double Sphere (DSCM) [33], where multiple parameter sets can describe the same physical camera. Such ambiguities[1] can result in treating different but correct parameter sets as errors, causing networks to learn parameter ranges instead of camera characteristics, diminishing accuracy and generalization.

Building on these insights, we propose a novel *two-step*, learning-based framework for radially symmetric camera calibration designed to overcome the limi-

[1] Refer to SM 7 for details about ambiguities and visualizations for EUCM and DSCM.

tations of traditional end-to-end approaches. Our framework takes an image $\mathcal{I}$ and a radially symmetric model $\mathcal{M}$ as inputs to estimate camera parameters according to $\mathcal{M}$. In the first step, a convolutional neural network (CNN) processes $\mathcal{I}$ and, rather than directly regressing camera parameters, estimates an implicit camera representation, the VaCR (*Viewing-angle Camera Representation*). Designed for radially symmetric cameras, the VaCR maps each image pixel into the ray of light projecting onto it, *i.e.*, the VaCR is independent of the input camera model $\mathcal{M}$. Secondly, a robust fitting procedure estimates the optimal parameters for $\mathcal{M}$ that best fit the estimated VaCR. Our approach, in Fig. 1, offers advantages over existing methods. **(i)** The model-agnostic VaCR allows considering $\mathcal{M}$ only in the robust fitting step, thus bypassing network retraining when the camera model changes. **(ii)** This eliminates the need for new model-specific data for network training when the camera model changes. **(iii)** By using the unambiguous VaCR as an intermediate camera representation, our robust fitting converges to a valid set of parameters for the target camera even when the input camera model has inherent ambiguities, *e.g.*, EUCM and DSCM.

In sum, we advance concerning the state-of-the-art with three contributions:

1. We introduce the VaCR, a camera representation for radially symmetric cameras devoid of ambiguities and independent of a specific camera model.
2. We design a VaCR estimation CNN, exploiting radial symmetries for improved efficiency and accuracy. By estimating the VaCR instead of camera parameters directly, our CNN is trained independently of any camera model and supports many cameras without retraining or architectural changes.
3. These elements contribute to a novel *two-step* calibration framework for radially symmetric cameras: first, the CNN predicts the VaCR; then, a robust fitting procedure minimizes a cost function based on the VaCR to estimate parameters specific to the input camera model, even for ambiguous models.

We validate our framework's efficacy using public datasets, showing that it surpasses current leading methods in camera calibration through quantitative and qualitative tests. Additionally, we attain results on par with or better than the state-of-the-art in image rectification despite it not being our main objective.

2 Problem Formulation

Calibration of radially symmetric cameras is framed as follows. The inputs are **(i)** an image $\mathcal{I}$ captured by the target camera, **(ii)** a radially symmetric camera model $\mathcal{M}$ depending on a set of intrinsic parameters $\mathbf{i}_\mathcal{M} = [f, \mathbf{d}]$, where $f \in \mathbb{R}$ is the focal length and $\mathbf{d} = [d_1, \ldots, d_n] \in \mathbb{R}^n$ is a set of coefficients to model the lens distortion. Our goal is to recover $\mathbf{i}_\mathcal{M}$ in a learning-based framework, where the training dataset comprises a set of distorted images, each labeled with the set of ground truth parameters that model the image distortion.

This work focuses on wide-angle cameras, *i.e.*, cameras with an angular field-of-view (AFOV) greater or equal to 80°. We follow the common assumptions of zero skew, principal point at the camera center, and square pixel aspect ratio.

3 Related Work

Camera calibration approaches fall into *geometric*-based and *learning*-based.

Geometric-based methods use calibration objects [32, 39], line detection [1, 3, 5, 11, 30, 38], or vanishing points [23, 27] to establish world-to-image relationships, enabling accurate and reliable camera parameter estimation. Despite their effectiveness, they struggle in unstructured environments without manual input.

Learning-based methods, based on CNNs, calibrate cameras end-to-end from a single image from an uncontrolled environment. These methods primarily target intrinsic and distortion parameter estimation, although recent works [7, 12, 28] have also addressed extrinsic parameter calibration. While our work is focused on intrinsics, it can be adapted to also estimate extrinsics, as discussed in Sec. 5.5.

We categorize learning-based methods into regression and reconstruction.
(i) Regression models [4, 13, 14, 17, 18, 24, 29, 34, 35] estimate parameters based on a predefined camera model $\mathcal{M}$ specific to each method. Despite effectiveness, they lack accuracy and generalization due to their *single-model* dependency, which hampers network adaptability to different camera types. This single-model setup also induces a narrow range of distortion parameters in the training datasets, yielding poor calibration accuracy when parameters are outside predefined ranges. While "general" camera models have been proposed [34], they lack the representational capacity to suit cameras with complex projection functions.

Liao et al. [20] model-free approach first uses adversarial learning for image rectification, estimating a parameter-free distortion distribution map (DDM) without relying on a predefined camera model. Yet, the DDM does not exploit radial symmetries and the overall method underperforms in accuracy and efficiency compared to the state-of-the-art, as shown in Sec. 5. The adversarial training process further introduces challenges in loss balancing between the generator and discriminator. In contrast, our method introduces the VaCR, a camera model-independent representation. Unlike the DDM, which assigns a distortion level to each pixel based on the 2D image coordinates' ratio between the distorted and the rectified image, the VaCR assigns each pixel to a bearing vector, describing the direction of the ray of light incident onto said pixel. Including 3D information in our camera representation aligns with recent advances in learning-based calibration, mainly through adopting bearing losses [24, 34] to drive the network training on re-projection errors.

Another limitation of successful, recent calibration methods, *e.g.*, Wakai et al. [34], lies in their reliance on loss functions formulated from the camera model's undistortion equations, which, to be differentiable, require the model's 2D-to-3D back-projection function to have a unique, closed-form solution for computing bearing vectors. However, sophisticated camera models like EUCM [15] and DSCM [33] do not fulfill this condition, restricting their use in these methods. In scenarios lacking a differentiable back-projection function, other approaches, such as direct loss computation from camera parameters [4] or iterative solutions for the back-projection function [24] have been proposed, yet these often fall short in accuracy. Moreover, computing the loss directly from camera parameters, *e.g.*,

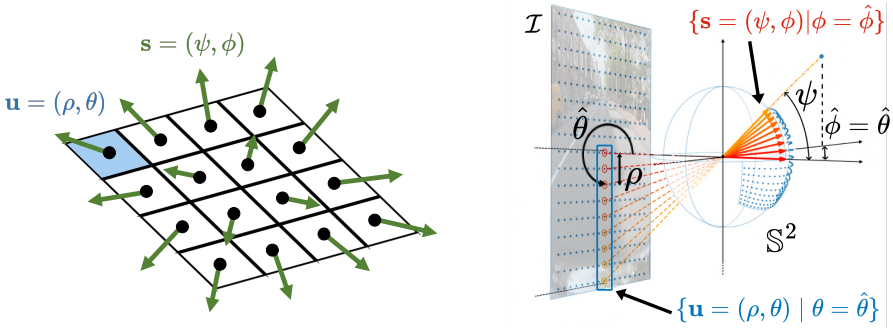

Fig. 2: Left. Mapping each image pixel $\mathbf{u} = (\rho, \theta)$ to a bearing vector $\mathbf{s} = (\psi, \phi)$. **Right.** For pixels $\mathbf{u}$ with the same angle $\theta = \hat{\theta} = \frac{3\pi}{4}$ (outlined in cyan), their bearing vectors (in shades of red) have constant $\phi = \hat{\phi} = \hat{\theta}$, but their ψ values vary by ρ. The VaCR comprises the set of angles ψ of the bearing vectors of all pixels with $\theta = \hat{\theta}$.

using the root-mean-square error, is problematic for models with ambiguities, where multiple parameter sets can represent the same camera. By estimating the VaCR rather than camera parameters directly, our approach sidesteps the requirement of a closed-form solution to the model's back-projection function, as the VaCR is model-independent and the loss used to train the VaCR estimation network does not involve any model's back-projection function.

(ii) Reconstruction models [6, 9, 19, 36] employ adversarial learning, leveraging multi-scale information to train generators and discriminators for image undistortion without explicitly estimating camera parameters. [40] employs polar coordinates within to estimate a 1D distorted-to-rectified flow instead of a 2D flow. Conversely, our two-step regression framework leverages polar coordinates to enable specialized pooling operations that exploit the radial symmetries of cameras. In Sec. 5.4 we compare reconstruction approaches to our method in image rectification, showing that we attain consistently superior results.

4 Method

This section outlines our two-step calibration method for radially symmetric cameras. Initially, in Sec. 4.1, we present our model-independent camera representation, the VaCR. Next, in Sections 4.2 - 4.3, we detail the CNN designed for VaCR estimation from the input image $\mathcal{I}$, which enables us to disentangle network training from a specific camera model. Lastly, we describe the robust fitting process in Sec. 4.4, fitting camera parameters for the input model $\mathcal{M}$ via nonlinear least squares to the estimated VaCR.

4.1 Viewing-angle Camera Representation

This section introduces the *Viewing-angle Camera Representation*, or VaCR, a key component of our calibration framework that enables full characterization of any radially symmetric camera, independently of a specific camera model.

Image rectification may be achieved by mapping each pixel $\mathbf{u}$ to the direction of its incident light ray $\mathbf{s}$. For instance, in the ideal case of a pinhole camera, where there is no distortion, and the projection equation writes $\lambda\mathbf{u} = K\,(R\,|\,t)\,\mathbf{U}$, it is sufficient to know the matrix of intrinsic parameters K to determine the direction of the light ray $\mathbf{s} = K^{-1}\mathbf{u}$ for a pixel $\mathbf{u}$. On the other hand, to account for lens distortion, it is necessary to model the camera's physics and employ analytical expressions to back-project a pixel in its optical ray. The idea behind the VaCR is to avoid using a complex analytical camera model. Since the number $H \times W$ of pixels in $\mathcal{I}$ is finite, the back-projection equations are reduced to a look-up table, which can be further compressed by exploiting common assumptions such as invariance to the radial symmetries in cameras.

We define the VaCR as a function $\pi : \mathcal{I} \to \mathbb{S}^2$ mapping each image point $\mathbf{u} \in \mathcal{I}$ to a bearing vector $\mathbf{s} = (\psi, \phi)$ on the unit sphere $\mathbb{S}^2$ (see Fig. 2-left). π is fully characterized by $H \times W \times 2$ parameters, as each bearing vector $\mathbf{s}$ comprises two angles, ψ and ϕ, describing the direction of the 3D ray projecting onto $\mathbf{u}$.

In the context of estimating the VaCR in a learning-based framework, we aim to further reduce the number of parameters defining π. By exploiting the angle-invariance property of radially symmetric cameras, illustrated in Fig. 2-right, we can accomplish this reduction without loss of information. Specifically, we express the image point $\mathbf{u} = (\rho, \theta)$ in polar coordinates, with ρ denoting the radial distance and θ the azimuth angle. Under the assumption of radial symmetry, any generic radial projection function does not alter the angle ϕ when projecting $\mathbf{s}$ onto $\mathbf{u}$, *i.e.*, $\phi = \theta$.[2] Knowing that $\phi = \theta$ and assuming θ is known for any image point, each bearing vector has only one undetermined parameter, *i.e.*, ψ, meaning the parameter count of π is reduced to $H \times W$.

We also observe that the radial distance ρ of a pixel depends only on the angle ψ of its bearing vector. Thus, all pixels at the same radial distance, $\rho = \hat{\rho}$, will map to bearing vectors with the same angle $\psi = \hat{\psi}$ (see Fig. 2-right). Hence, it is unnecessary to map each pixel to an angle ψ; instead, encoding ψ for only one representative pixel per radial distance value ρ suffices. Consequently, the function π simplifies to N parameters, where N is the count of discrete radial distances in $\mathcal{I}$, effectively reducing to $N = \frac{H}{2} = \frac{W}{2}$ for square images.

The formulation of the VaCR, optimized for radial symmetry, is:

$$\text{VaCR} : \{\rho_i\}_{i=1}^{N} \to \mathbb{R}\,, \tag{1}$$

mapping each radial distance to the angle ψ of the corresponding bearing vector. This definition highlights that the VaCR is unambiguous, as mapping each radial distance to an angle ψ ensures that, if two cameras differ, their VaCRs will not match due to their distinct back-projection functions yielding different angles ψ.

4.2 VaCR Estimation Network

We detail the architecture of the proposed VaCR estimation CNN, summarized in Fig. 3. Unlike existing approaches, we first transform the input image

[2] The verification of the angle-invariance property depends on the adopted camera model; we provide a proof for the DSCM model in SM 8.

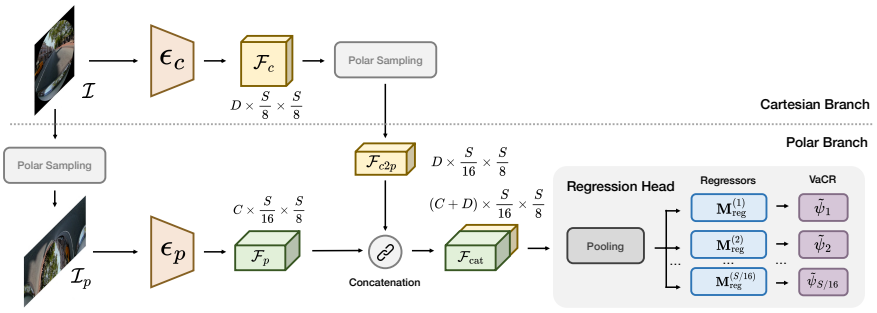

Fig. 3: VaCR Estimation Network: Dual encoders ϵ_p and ϵ_c derive features from the polar and Cartesian images. Cartesian features ($\mathcal{F}_c$) are first mapped to polar coordinates ($\mathcal{F}_{c2p}$) and then concatenated with polar features ($\mathcal{F}_p$), forming $\mathcal{F}_{\text{cat}}$. After pooling operations on $\mathcal{F}_{\text{cat}}$, distinct regressors $\mathbf{M}_{\text{reg}}^{(i)}$ estimate the VaCR. We provide an expanded view of the regression head in Fig. 4.

$\mathcal{I}$ into polar coordinates $\mathcal{I}_p$ before providing it to the feature encoder. This transformation reorganizes image data, enabling specialized pooling operations that regularize the VaCR estimation while also reducing the network parameter count, improving efficiency. To address information losses from the Cartesian-to-polar conversion, we employ an additional encoder that extracts features from the Cartesian image $\mathcal{I}$, which are then concatenated with polar features in later CNN stages, creating a specialized feature volume for VaCR regression.

Image pre-processing. The input image $\mathcal{I}$ is zero-padded to form a square image of size S and, then, is then transformed into polar coordinates, resulting in $\mathcal{I}_p$. The height of $\mathcal{I}_p$ is $\frac{S}{2}$, matching the inscribed circle's radius, and width W_p is set based on the chosen angular resolution, which is thus a parameter that can be tuned. By setting $W_p = S$, we balance between reducing information loss and minimizing pixel deformation.[3]

Feature extraction. Our network employs two distinct ConvNeXt [22] encoders, ϵ_p and ϵ_c, which separately extract features from $\mathcal{I}_p$ and $\mathcal{I}$ respectively. Starting from $\mathcal{I}_p$, sized $\frac{S}{2} \times S$, ϵ_p produces a feature volume $\mathcal{F}_p$ of dimensions $C \times \frac{S}{2k} \times \frac{S}{k}$, where C denotes the number of feature channels and k the downsampling factor of the feature encoder. Similarly, ϵ_c produces a feature volume $\mathcal{F}_c$, sized $D \times \frac{S}{k} \times \frac{S}{k}$. As discussed in Sec. 5.5, we opt to extract fewer Cartesian feature channels D compared to polar ones C due to observed diminishing improvements in VaCR accuracy when incrementing D beyond a certain threshold.

Combining features. Features $\mathcal{F}_p$ and $\mathcal{F}_c$ merge into $\mathcal{F}_{\text{cat}}$ for VaCR regression. To achieve this, $\mathcal{F}_c$ is transformed into polar coordinates as $\mathcal{F}_p$, creating $\mathcal{F}_{c2p}$ sized $D \times \frac{S}{2k} \times \frac{S}{k}$. This process parallels the transformation of $\mathcal{I}$ into $\mathcal{I}_p$, but, notably, it is performed after image features are extracted from $\mathcal{I}$ by a specialized encoder. Thus, $\mathcal{F}_{c2p}$ comprises features extracted from the Cartesian image, with

[3] Refer to SM 9 for an ablation study on W_p's effects on VaCR regression.

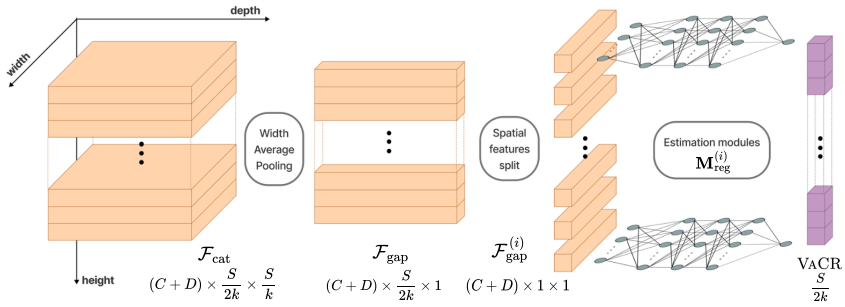

Fig. 4: Regression Head of VACR Network. The feature volume $\mathcal{F}_{\text{cat}}$ undergoes width-wise averaging to form $\mathcal{F}_{\text{gap}}$, which is then divided height-wise into slices $\mathcal{F}_{\text{gap}}^{(i)}$. Independent estimation modules $\mathbf{M}_{\text{reg}}^{(i)}$ predict the VACR, i.e., $\frac{S}{2k}$ viewing angles.

a spatial arrangement compatible to that of polar features $\mathcal{F}_p$. By concatenating $\mathcal{F}_p$ and $\mathcal{F}_{c2p}$ along the channel dimension, we obtain a combined feature volume $\mathcal{F}_{\text{cat}}$ sized $(C+D) \times \frac{S}{2k} \times \frac{S}{k}$ having a consistent spatial arrangement throughout.

VACR Regression. In the network's final stage, the VACR is regressed from the combined feature volume $\mathcal{F}_{\text{cat}}$. The polar transformation brings points at the same radial distance, or viewing angles in the VACR, along the same row in the polar image $\mathcal{I}_p$. This setup ensures that viewing angles in the VACR for adjacent image rows are highly correlated due to their similar radial distances. Capitalizing on this spatial correlation, the kernels of our CNN traverse the axes of the polar image, processing image regions that are highly correlated for VACR regression, even when the receptive field is compact.

We design the regression head to exploit the same-row arrangement in $\mathcal{I}_p$ of points at equal radial distance. Specifically, we employ Global Average Pooling across the width of $\mathcal{F}_{\text{cat}}$ to obtain a feature volume $\mathcal{F}_{\text{gap}}$, sized $C \times \frac{S}{2k} \times 1$ (see Fig. 4). This strategy averages features at each row, regularizing information for each radial distance, *i.e.*, viewing angle in the VACR, while reducing the size of the feature volume. Notably, due to feature downsampling, each row in $\mathcal{F}_{\text{cat}}$ combines data from k radial distances, as the image is downsampled from size $\frac{S}{2}$ to $\frac{S}{2k}$. This condensation increases efficiency while maintaining accuracy since adjacent radial distances are usually mapped to nearly identical viewing angles.[4]

We exploit the spatial relationship between $\mathcal{F}_{\text{gap}}$ and the VACR, where each row in $\mathcal{F}_{\text{gap}}$ maps to an angle in the VACR, and divide $\mathcal{F}_{\text{gap}}$ along its height into feature slices $\{\mathcal{F}_{\text{gap}}^{(i)}\}_{i=1}^{S/2k}$, each sized $C \times 1 \times 1$. Each slice feeds into a separate estimation module $\mathbf{M}_{\text{reg}}^{(i)}$, comprising fully connected layers that regress a viewing angle (see Fig. 4). This regression head architecture, with dedicated $\mathbf{M}_{\text{reg}}^{(i)}$ with distinct weights for each slice, exploits the spatial feature arrangement in $\mathcal{F}_{\text{gap}}$, providing individual regression pathways for each angle. Distinct statistical properties of features at different viewing angles justify the use of indi-

[4] SM 9 analyzes the effects of different downsampling rates on calibration accuracy.

vidual $\mathbf{M}_{\text{reg}}^{(i)}$ modules and allows for tailored specialization, contrary to existing methods that regress camera parameters from a unified feature set.

4.3 Network Training

We train the feature extractors, ϵ_p and ϵ_c, and the regression modules $\mathbf{M}_{\text{reg}}^{(i)}$ end-to-end. In each mini-batch of size B, for each sample j, the CNN estimates the VaCR $\{\rho_i \rightarrow \tilde{\psi}_{ij}\}_{i=1}^{S/2k}$, comprising $\frac{S}{2k}$ radial distance ρ_i and angle estimate $\tilde{\psi}_{ij}$ pairs. The loss $\mathcal{L}$ is:

$$\mathcal{L}(\tilde{\psi}_{ij}) = \frac{1}{B\frac{S}{2k}} \sum_{j=1}^{B} \sum_{i=1}^{S/2k} \text{Huber}(\|\rho_i - P_{r,j}(\tilde{\psi}_{ij}, \mathbf{i}_{\text{gt},j})\|_2) \,, \tag{2}$$

where $P_{r,j}$ is the radial projection for the j-th sample, and $\mathbf{i}_{\text{gt},j}$ the ground truth camera parameters. The predicted viewing angles, $\tilde{\psi}_{ij}$, are used within P_j along with ground truth parameters $\mathbf{i}_{\text{gt},j}$ to evaluate the Euclidean distance from the correct ρ_i. This loss computation is independent on any specific camera model, allowing each batch sample to contribute its own model with projection function P_j and parameters $\mathbf{i}_{\text{gt},j}$. The Huber($\bullet$) denotes the Huber loss.

Dataset Generation. We train the network on datasets comprising synthetically distorted images labeled with ground truth camera parameters. Existing approaches [4, 24, 34] generate synthetic datasets using a fixed camera model by uniformly sampling the focal length f and distortion parameters within a preset range. This strategy leads to a skewed distribution of the angular field-of-view (AFOV) in the dataset when the camera model has ambiguities – different parameters describe the same camera. Since AFOV is crucial for camera characterization, having a skewed distribution may induce biases in the network and limit its ability to generalize to certain camera types.

Our setup synthesizes images using DSCM [33] and EUCM [15] camera models, as they exhibit superior representational capabilities, as discussed in SM 7. Unlike existing approaches, we directly sample the AFOV between 80° and 190° before sampling model-specific parameters, ensuring uniform AFOV distribution in the dataset. Then, for DSCM, parameters α and ϵ are uniformly sampled in $[0, 1]$ and $[-1, 1]$, respectively, and the focal length is derived as $f = S/\left[2\mathcal{D}\left(\frac{1}{2}\text{AFOV}, a, \xi\right)\right]$, where S is the image size and $\mathcal{D}(\theta, a, \epsilon)$ a function of AFOV, a, ϵ. EUCM settings follow similarly, as detailed in SM 7.

4.4 Fitting Camera Parameters to the VaCR

The final step involves estimating the intrinsic parameters $\mathbf{i}_{\mathcal{M}}$ for the input camera model $\mathcal{M}$ from the VaCR returned from our estimation network (Sec. 4.2), *i.e.*, $\{\rho_i \rightarrow \tilde{\psi}_i\}_{i=1}^{S/2k}$. Parameters for the input model $\mathcal{M}$ are obtained by minimizing the following objective function for $\mathbf{i}_{\mathcal{M}}$:

$$\arg\min_{\mathbf{i}_{\mathcal{M}}} \sum_{i=1}^{S/2k} \alpha\left[\left(\rho_i - P_{r,\mathcal{M}}(\tilde{\psi}_i, \mathbf{i}_{\mathcal{M}})\right)^2\right] , \tag{3}$$

where $P_{r,\mathcal{M}}$ is the input model $\mathcal{M}$'s radial projection function that depends on the predicted viewing angle $\tilde{\psi}_i$ for radial distance ρ_i and parameters $\mathbf{i}_{\mathcal{M}}$. α is the Cauchy loss function $\alpha(s) = \ln(1+s)$, which reduces the impact of outliers.

To address this optimization, we apply the *trust region reflective* (TRF) algorithm. TRF optimizes an initial guess $\mathbf{i}^0_{\mathcal{M}}$ of the parameters in a constrained parameter space, which we define based on prior knowledge about the input camera model $\mathcal{M}$ to improve the robustness and computational speed of the algorithm. The initial estimate $\mathbf{i}^0_{\mathcal{M}}$ of the camera parameters is obtained by the stochastic evolution method in [31].

5 Experiments

We evaluate our method and compare it to state-of-the-art approaches for the tasks of: **(i)** camera calibration (Sec. 5.3), **(ii)** image rectification (Sec. 5.4), as the quality of rectified images provides a good qualitative and quantitative indicator of calibration accuracy. The public datasets used to generate our synthetic datasets for training and testing are presented in Sec 5.1, while we provide implementation details in Sec 5.2. Finally, we conduct an ablation study (Sec. 5.5) to demonstrate the effectiveness of select design choices in the proposed framework.

5.1 Datasets

We consider the following publicly available datasets for synthetic data generation: **(i)** KITTI-360 [21] provides urban and landscape images from wide-angle fisheye cameras with AFOV over 190°. **(ii)** StreetLearn [26] offers 360° urban panoramas from Google Street View, rich in lines, arcs, and repeated geometric patterns. **(iii)** SILDa [2] includes images from a low-end spherical camera with 200° AFOV, capturing diverse urban outdoor conditions over a year. **(iv)** WoodScape [37] contributes a rich automotive dataset with images from various 190° AFOV fisheye cameras and different aspect ratios.

Images are generated at a resolution of 400x400 pixels, cropping non-square originals to fit this size. We partition the starting datasets into training and testing sets before data generation, ensuring no overlap. This process yields 145953 training samples and 14082 testing samples across all datasets.

5.2 Implementation Details

In our implementation, two ConvNexT [22] feature extractors, ϵ_p and ϵ_c, are utilized, with initial pre-trained weights from ImageNet [8] and downsampling factor $k = 8$. Specifically, the weights of ϵ_p and ϵ_c are frozen for two training epochs, then unfrozen for end-to-end training. Training data augmentation includes rotation and flipping. The learning rate is set at 10^{-4} for 100 epochs, with a batch size of 40. An early stopping mechanism is applied if there is no improvement after 10 consecutive epochs, starting after the first 10 epochs. We employ the AdamW [25] optimizer with a learning rate of 10^{-4}, weight decay of

	KITTI-360 [21]				StreetLearn [26]				SILDa [2]				WoodScape [37]			
Method	f [px]	a	k_1	Re [px]	f [px]	a	k_1	Re [px]	f [px]	a	k_1	Re [px]	f [px]	a	k_1	Re [px]
López-Antequera [24]	54.601	-	-	28.401	52.163	-	-	32.044	48.170	-	-	28.691	18.210	-	-	22.384
DeepCalib [4]	48.082	0.280	-	21.936	37.033	0.219	-	19.692	33.639	0.211	-	14.905	11.041	0.105	-	12.339
Ours w/ UCM	11.790	**0.062**	-	5.272	14.482	**0.070**	-	6.950	**11.302**	**0.074**	-	5.376	7.069	**0.065**	-	3.350
Wakai [34]	18.035	-	0.049	7.882	17.501	-	0.051	8.925	14.538	-	0.059	8.060	9.263	-	0.054	5.017
Ours w/ KB3	**11.623**	-	**0.030**	5.324	**14.217**	-	**0.037**	6.861	11.441	-	**0.042**	5.361	**6.731**	-	**0.031**	3.316
Liao et al. [20]	-	-	-	6.021	-	-	-	7.956	-	-	-	8.405	-	-	-	6.899
Ours w/ EUCM	-	-	-	4.633	-	-	-	6.861	-	-	-	5.197	-	-	-	**3.239**
Ours w/ DSCM	-	-	-	**4.015**	-	-	-	**6.676**	-	-	-	**5.097**	-	-	-	3.265

Table 1: Calibration Evaluation. Absolute parameter errors and reprojection errors for the datasets in Sec. 5.1. Absences in the table imply unavailability or inapplicability of certain parameters for respective methods. For all metrics, lower is better.

5×10^{-3}, and β parameters of (0.9, 0.999). Training is conducted on an NVIDIA RTX 4090, allowing the network to converge in about 23 hours.

5.3 Evaluation of Camera Calibration

Evaluation Metrics. We evaluate parameter prediction accuracy using Mean Absolute Error (MAE) and Mean Reprojection Error (Re), with a focus on Re for camera models with ambiguities like EUCM and DSCM. To compute Re, we back-project 160,000 points per test image onto the unit sphere at the camera's origin using the model's back-projection function with ground truth parameters. Subsequently, we reproject these points from the sphere onto the image using the predicted parameters. Re is then assessed as the average distance between the original and reprojected image points (refer to SM 10 for a detailed definition of error metrics.)

Comparing Methods. We compare our method with state-of-the-art learning-based calibration techniques: DeepCalib [4], López-Antequera [24], and Wakai [34], each setting the standard for different camera models. DeepCalib is based on UCM [10], while López-Antequera and Wakai use 4th and 3rd-order polynomial projection functions, respectively, with Wakai's model known as KB3. We also include Liao et al.'s model-free approach [20], which, despite being parameter-free, allows for Re computation from the predicted distortion distribution map (DDM). As López-Antequera, Wakai, and Liao et al. are not publicly available, we replicated their implementation and trained them on our datasets.

Quantitative Results. Tab. 1 shows our method outperforming competitors in camera calibration. With the UCM model, our method surpasses DeepCalib [4] in Re and MAE across all datasets. With KB3, we are better than Wakai [34] in all metrics, though with a smaller margin than against DeepCalib. Lopez [24] performs the weakest, particularly with large AFOV cameras, as confirmed by [34]. Liao et al.'s [20] model-free approach, which doesn't provide direct parameter

	KITTI-360 [21]			StreetLearn [26]			SILDa [2]			WoodScape [37]		
Methods	SSIM ↑	PSNR ↑	FID ↓	SSIM ↑	PSNR ↑	FID ↓	SSIM ↑	PSNR ↑	FID ↓	SSIM ↑	PSNR ↑	FID ↓
Auto-DE [5]	0.179	3.48	285.1	0.146	4.21	210.1	0.140	4.01	239.0	0.137	3.20	346.1
Auto-DC [1]	0.186	3.72	260.4	0.136	4.78	195.3	0.154	4.22	218.2	0.150	3.54	306.7
DeepCalib [4]	0.517	14.39	97.0	0.452	16.02	84.0	0.596	18.84	79.2	0.501	17.41	86.9
López-Antequera [24]	0.487	13.92	110.4	0.439	15.18	97.3	0.391	12.91	137.0	0.429	15.37	110.4
Blind [16]	0.534	14.76	94.5	0.490	17.13	89.4	0.561	17.48	83.0	0.490	17.38	87.3
Liao et al. [20]	0.520	14.60	96.1	0.448	16.09	91.7	0.559	17.51	82.8	0.502	17.28	87.0
Wakai et al. [34]	0.521	14.89	95.0	0.473	16.21	90.4	0.552	17.21	84.4	0.508	17.35	86.8
Jin et al. [14]	0.501	14.11	99.6	0.432	15.20	95.2	0.511	16.28	93.4	0.411	15.03	110.4
DR-GAN [19]	0.491	14.02	102.4	0.429	14.96	100.6	0.501	15.98	98.6	0.403	14.20	134.2
PCN [36]	0.504	14.39	96.4	0.435	15.77	91.2	0.533	17.10	89.4	0.429	16.22	93.3
Ours w/ UCM	0.582	16.35	84.6	0.518	17.42	86.2	0.601	18.69	80.1	0.524	17.89	83.5
Ours w/ KB3	0.581	16.28	85.4	0.516	17.35	88.6	0.602	18.75	79.6	0.518	17.54	84.0
Ours w/ EUCM	**0.610**	**17.53**	**78.4**	0.520	17.60	81.8	0.609	19.01	76.5	**0.564**	**18.95**	**80.5**
Ours w/ DSCM	0.583	16.37	83.2	**0.523**	**17.77**	**80.4**	**0.612**	**19.06**	**73.7**	0.544	18.54	81.2

Table 2: Image Rectification Evaluation. Our method versus state-of-the-art across datasets in Sec. 5.1, with "↑" indicating better when higher, "↓" when lower.

estimates, shows inferior *Re* compared to ours. For the ambiguous EUCM and DSCM, we omit MAE due to the high errors when calibrations are equivalent yet defined by different sets of parameters. Using DSCM or EUCM models, our method achieves the best *Re* across all datasets. These results prove that our method can effectively adapt to several camera models, even those ambiguous but with high representational power.

5.4 Evaluation of Image Rectification

Evaluation Metrics. Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) evaluate image quality, with PSNR focusing on detail accuracy and SSIM on structural integrity. Additionally, Fréchet Inception Distance (FID), leveraging Wasserstein-2 distance, assesses distributional discrepancies. We use these metrics for objective assessment of our experimental results.

Comparing Methods. Our method is compared with state-of-the-art geometric and learning-based models. In geometric approaches, we include Auto-DE [5] and Auto-DC [1]. For learning-based, we consider both regression and reconstruction methods: DeepCalib, López-Antequera, Liao et al., Wakai, Jin et al. [14] plus Blind [16], designed specifically for image rectification. For reconstruction approaches, focusing on rectification without parameter estimation, we evaluate DR-GAN [19] and PCN [36], the current top performers.

Quantitative Comparison Results. Tab. 2 demonstrates the superior image rectification performance of our method over state-of-the-art geometric and learning-based approaches. Notably, using the EUCM and DSCM models, we achieve the highest SSIM and PSNR and lowest FID, demonstrating improvements in image quality and structural integrity. Blind [16] encounters issues

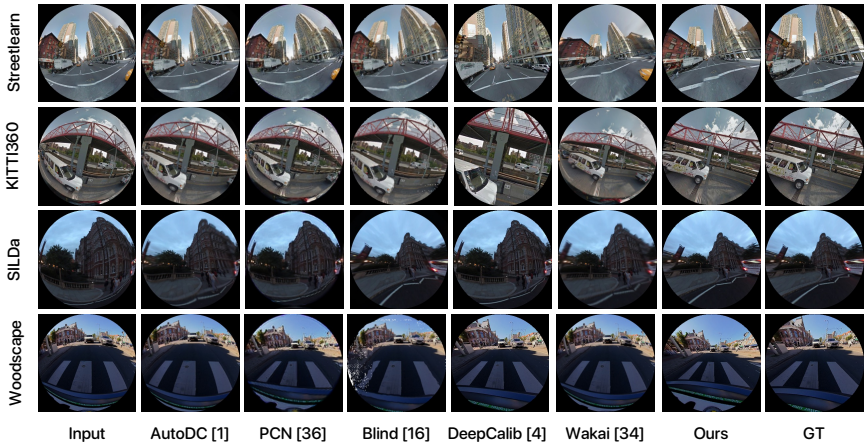

Fig. 5: Image Rectification Qualitative Results. Comparisons to state-of-the-art methods illustrate the effectiveness of our technique across different datasets.

with strong image distortions, while DeepCalib [4] and Wakai [34] improve central regions, but exhibit issues near the boundaries. López-Antequera [24] faces challenges with large AFOV, a limitation known from [34]. DR-GAN [19] yields blurred images with artefacts, while Liao et al. [20] and PCN [36] present degraded quality in localized areas. Jin et al. [14] yields lower scores than all methods except [24] due to using only the FOV parameter for modeling distortions. Our approach excels particularly when using EUCM and DSCM, demonstrating superior overall results. The model-free Liao et al. [20] is our closest competitor, but our approach surpasses it in all metrics, confirming our claims of accuracy and adaptability to various camera models.

5.5 Ablation Study & Discussion

We conduct ablations to verify the effectiveness of core settings and components.

Qualitative Comparison is performed against competitors [4, 16, 36] on the synthetic test set. In Fig. 5 it can be appreciated that our method's rectified images exhibit less distortion.

Impact of Polar vs. Cartesian. Tab. 3 illustrates the impact of varying the number of feature channels extracted from the polar (C) and Cartesian image (D) respectively, confirming that the proposed CNN can effectively exploit the polar image layout to produce accurate VACR estimates. By increasing D with constant C, we observe diminishing improvements in image rectification accuracy after the $D = 96$ threshold. Conversely, an increase in C with constant D usually yields non-negligible improvements in accuracy. We set $(C, D) = (384, 96)$ achieving an optimal trade-off between accuracy and efficiency. Notably, eliminating polar features ($C = 0$) and relying on Cartesian features only severely impacts performance, leading to worse results than most methods in Tab. 2.

C/D	**0**	**96**	**192**	**384**
0	-	0.289/10.61	0.294/10.78	0.299/10.84
96	0.508/14.05	0.512/14.21	0.515/14.29	0.517/14.33
192	0.569/15.97	0.571/16.00	0.572/16.03	0.572/16.04
384	0.594/16.74	0.610/17.53	0.611/17.55	0.611/17.56

Table 3: Comparison of PSNR and SSIM in KITTI-360 [21] for image feature channels in polar (C) and Cartesian (D) coordinates. Each cell represents the metrics formatted as PSNR/SSIM. The highlighted cell represents the configuration used in our experiments.

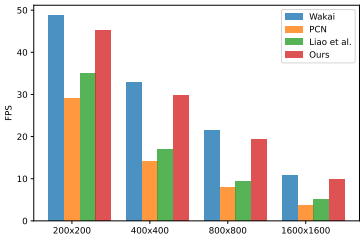

Fig. 6: Running FPS comparison between our framework and Wakai [34], PCN [36], Liao et al. [20] on processing different resolution images (x-axis).

Running Times. We benchmark our method against Wakai [34], PCN [36], and Liao et al. [20] across different image resolutions. Results in Fig. 6 show our method achieves higher FPS (frames per second) than both PCN and Liao et al., supporting our claims of efficiency and accuracy. The polar transformation detailed in Sec. 4.2 helps reduce the parameter count by enabling width average pooling, which compacts feature volumes and reduces parameters in subsequent network layers. Our network uses 6.6M parameters, significantly fewer than the 39M parameters of our model-free competitor, Liao et al. [20]. Although Wakai, which regresses camera parameters directly without a robust fitting procedure, is up to 13% faster, it is less accurate, as demonstrated in Tab. 1 and Tab. 2.

Extrinsic parameter estimation. While our focus is intrinsic parameter estimation, our method can extend to extrinsic estimates, like camera tilt and roll, by including additional regressors, as proposed in [34]. In SM 9, we demonstrate our approach's comparable state-of-the-art performance in these extensions.

6 Conclusion

Driven by the need for accurate and efficient calibration of wide-angle radially symmetric cameras, we introduce a novel two-step learning-based framework overcoming many of the limitations of existing methodologies. Our approach excels in calibrating radially symmetric cameras and rectifying images, representing a significant leap forward in this domain.

Acknowledgements: This paper is supported by PNRR-PE-AI FAIR project funded by the NextGeneration EU program and by GEOPRIDE ID: 2022245ZYB, CUP: D53D23008370001 (PRIN 2022 M4.C2.1.1 Investment).

References

1. Alemán-Flores, M., Alvarez, L., Gomez, L., Santana-Cedrés, D.: Automatic lens distortion correction using one-parameter division models. Image Processing On Line **4**, 327–343 (2014) 1, 4, 12
2. Balntas, V.: SILDa: A Multi-Task Dataset for Evaluating Visual Localization. Medium (Apr 2019), https://medium.com/scape-technologies/silda-a-multi-task-dataset-for-evaluating-visual-localization-7fc6c2c56c74 10, 11, 12
3. Benligiray, B., Topal, C.: Blind rectification of radial distortion by line straightness. In: 2016 24th European Signal Processing Conference (EUSIPCO). pp. 938–942. IEEE (2016) 1, 4
4. Bogdan, O., Eckstein, V., Rameau, F., Bazin, J.C.: Deepcalib: A deep learning approach for automatic intrinsic calibration of wide field-of-view cameras. In: Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production. pp. 1–10 (2018) 1, 2, 4, 9, 11, 12, 13
5. Bukhari, F., Dailey, M.N.: Automatic radial distortion estimation from a single image. Journal of mathematical imaging and vision **45**, 31–45 (2013) 1, 4, 12
6. Chao, C.H., Hsu, P.L., Lee, H.Y., Wang, Y.C.F.: Self-supervised deep learning for fisheye image rectification. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 2248–2252. IEEE (2020) 5
7. Chen, Y., Schmid, C., Sminchisescu, C.: Self-supervised learning with geometric constraints in monocular video: Connecting flow, depth, and camera. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7063–7072 (2019) 4
8. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009) 10
9. Feng, H., Wang, W., Deng, J., Zhou, W., Li, L., Li, H.: Simfir: A simple framework for fisheye image rectification with self-supervised representation learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12418–12427 (2023) 5
10. Geyer, C., Daniilidis, K.: A unifying theory for central panoramic systems and practical implications. In: Computer Vision—ECCV 2000: 6th European Conference on Computer Vision Dublin, Ireland, June 26–July 1, 2000 Proceedings, Part II 6. pp. 445–461. Springer (2000) 11
11. Gonzalez-Aguilera, D., Gomez-Lahoz, J., Rodríguez-Gonzálvez, P.: An automatic approach for radial lens distortion correction from a single image. IEEE Sensors journal **11**(4), 956–965 (2010) 1, 4
12. Gordon, A., Li, H., Jonschkowski, R., Angelova, A.: Depth from videos in the wild: Unsupervised monocular depth learning from unknown cameras. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8977–8986 (2019) 4
13. Hosono, M., Simo-Serra, E., Sonoda, T.: Self-supervised deep fisheye image rectification approach using coordinate relations. In: 2021 17th International Conference on Machine Vision and Applications (MVA). pp. 1–5. IEEE (2021) 1, 4
14. Jin, L., Zhang, J., Hold-Geoffroy, Y., Wang, O., Blackburn-Matzen, K., Sticha, M., Fouhey, D.F.: Perspective fields for single image camera calibration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17307–17316 (2023) 1, 4, 12, 13

15. Khomutenko, B., Garcia, G., Martinet, P.: An enhanced unified camera model. IEEE Robotics and Automation Letters **1**(1), 137–144 (2015) 2, 4, 9
16. Li, X., Zhang, B., Sander, P.V., Liao, J.: Blind geometric distortion correction on images through deep learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4855–4864 (2019) 12, 13
17. Liao, K., Lin, C., Liao, L., Zhao, Y., Lin, W.: Multi-level curriculum for training a distortion-aware barrel distortion rectification model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4389–4398 (2021) 1, 4
18. Liao, K., Lin, C., Wei, Y., Li, F., Yang, S., Zhao, Y.: Towards complete scene and regular shape for distortion rectification by curve-aware extrapolation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14569–14578 (2021) 1, 4
19. Liao, K., Lin, C., Zhao, Y., Gabbouj, M.: Dr-gan: Automatic radial distortion rectification using conditional gan in real-time. IEEE Transactions on Circuits and Systems for Video Technology **30**(3), 725–733 (2019) 5, 12, 13
20. Liao, K., Lin, C., Zhao, Y., Xu, M.: Model-free distortion rectification framework bridged by distortion distribution map. IEEE Transactions on Image Processing **29**, 3707–3718 (2020) 1, 4, 11, 12, 13, 14
21. Liao, Y., Xie, J., Geiger, A.: Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(3), 3292–3310 (2022) 10, 11, 12, 14
22. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11976–11986 (2022) 7, 10
23. Lochman, Y., Dobosevych, O., Hryniv, R., Pritts, J.: Minimal solvers for single-view lens-distorted camera auto-calibration. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 2887–2896 (2021) 1, 4
24. Lopez, M., Mari, R., Gargallo, P., Kuang, Y., Gonzalez-Jimenez, J., Haro, G.: Deep single image camera calibration with radial distortion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11817–11825 (2019) 1, 2, 4, 9, 11, 12, 13
25. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017) 10
26. Mirowski, P., Banki-Horvath, A., Anderson, K., Teplyashin, D., Hermann, K.M., Malinowski, M., Grimes, M.K., Simonyan, K., Kavukcuoglu, K., Zisserman, A., et al.: The streetlearn environment and dataset. arXiv preprint arXiv:1903.01292 (2019) 10, 11, 12
27. Pritts, J., Kukelova, Z., Larsson, V., Chum, O.: Radially-distorted conjugate translations. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1993–2001 (2018) 1, 4
28. Ren, L., Song, Y., Lu, J., Zhou, J.: Spatial geometric reasoning for room layout estimation via deep reinforcement learning. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16. pp. 550–565. Springer (2020) 4
29. Rong, J., Huang, S., Shang, Z., Ying, X.: Radial lens distortion correction using convolutional neural networks trained with synthesized images. In: Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part III 13. pp. 35–49. Springer (2017) 1, 4

30. Santana-Cedrés, D., Gomez, L., Alemán-Flores, M., Salgado, A., Esclarín, J., Mazorra, L., Alvarez, L.: An iterative optimization algorithm for lens distortion correction using two-parameter models. Image Processing On Line **6**, 326–364 (2016) 1, 4
31. Storn, R., Price, K.: Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of global optimization **11**, 341–359 (1997) 10
32. Tsai, R.: A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. IEEE Journal on Robotics and Automation **3**(4), 323–344 (1987) 1, 4
33. Usenko, V., Demmel, N., Cremers, D.: The double sphere camera model. In: 2018 International Conference on 3D Vision (3DV). pp. 552–560. IEEE (2018) 2, 4, 9
34. Wakai, N., Sato, S., Ishii, Y., Yamashita, T.: Rethinking generic camera models for deep single image camera calibration to recover rotation and fisheye distortion. In: European Conference on Computer Vision. pp. 679–698. Springer (2022) 1, 2, 4, 9, 11, 12, 13, 14
35. Xue, Z., Xue, N., Xia, G.S., Shen, W.: Learning to calibrate straight lines for fisheye image rectification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1643–1651 (2019) 1, 4
36. Yang, S., Lin, C., Liao, K., Zhang, C., Zhao, Y.: Progressively complementary network for fisheye image rectification using appearance flow. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6348–6357 (2021) 5, 12, 13, 14
37. Yogamani, S., Hughes, C., Horgan, J., Sistu, G., Varley, P., O'Dea, D., Uricár, M., Milz, S., Simon, M., Amende, K., et al.: Woodscape: A multi-task, multi-camera fisheye dataset for autonomous driving. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9308–9318 (2019) 10, 11, 12
38. Zhang, M., Yao, J., Xia, M., Li, K., Zhang, Y., Liu, Y.: Line-based multi-label energy optimization for fisheye image rectification and calibration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4137–4145 (2015) 1, 4
39. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on pattern analysis and machine intelligence **22**(11), 1330–1334 (2000) 1, 4
40. Zhao, K., Lin, C., Liao, K., Yang, S., Zhao, Y.: Revisiting radial distortion rectification in polar-coordinates: A new and efficient learning perspective. IEEE Transactions on Circuits and Systems for Video Technology **32**(6), 3552–3560 (2021) 5