# L2R GAN: LiDAR-to-Radar Translation

Leichen Wang[1,2], Bastian Goldluecke[2], and Carsten Anklam[1]

[1] Research and Development of Radar Sensor, Daimler AG, Sindelfingen, Germany
leichen.wang, carsten.anklam@daimler.com
[2] Department of Computer and Information Science, University Konstanz, Germany
bastian.goldluecke@uni-konstanz.de

**Abstract.** The lack of annotated public radar datasets causes difficulties for research in environmental perception from radar observations. In this paper, we propose a novel neural network based framework which we call L2R GAN to generate the radar spectrum of natural scenes from a given LiDAR point cloud.
We adapt ideas from existing image-to-image translation GAN frameworks, which we investigate as a baseline for translating radar spectra image from a given LiDAR bird's eye view (BEV). However, for our application, we identify several shortcomings of existing approaches. As a remedy, we learn radar data generation with an occupancy-grid-mask as a guidance, and further design a set of local region generators and discriminator networks. This allows our L2R GAN to combine the advantages of global image features and local region detail, and not only learn the cross-modal relations between LiDAR and radar in large scale, but also refine details in small scale.
Qualitative and quantitative comparison show that L2R GAN outperforms previous GAN architectures with respect to details by a large margin. A L2R-GAN-based GUI also allows users to define and generate radar data of special emergency scenarios to test corresponding ADAS applications such as Pedestrian Collision Warning (PCW).

## 1 Introduction

In the past years, environmental perception based on cameras and Light Detection and Ranging (LiDARs) has made significant progress by using deep learning techniques. The basic idea is to design and train a deep neural network by feeding quantities of annotated samples. The training process enables the networks to effectively learn a hierarchical representation of pixels or points using high-level semantic features.

In contrast to LiDARs and camera, Frequency-Modulated Continuous-Wave (FMCW) radar operates at longer ranges and is substantially more robust to adverse weather and lighting conditions. Besides, on account of its compact size and reasonable price, radar is becoming the most reliable and most widely used sensor in Advanced Driver Assistance Systems (ADAS) applications. However, the research on deep learning for analyzing radar signals is still at a very early stage [1,2,3,4,5,6].

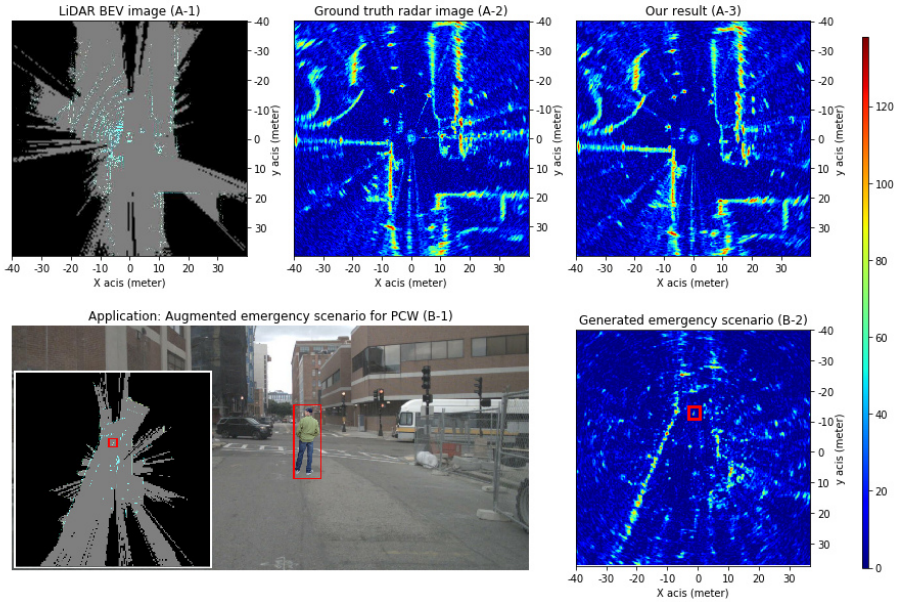


**Fig. 1.** A: We propose the L2R GAN for synthesizing radar spectrum images from given LiDAR point clouds. A-1: Input LiDAR BEV image with corresponding occupancy grid mask (black is unknown area, gray is free area). A-2: ground truth radar spectrum. A-3: generated radar spectrum of L2R GAN. B: An L2R-GAN-based GUI allows to define and generate the radar data of emergency scenarios to test corresponding ADAS application such as Pedestrian Collision Warning (PCW). B-1: example of an augmented emergency scenario for PCW in a camera and LiDAR BEV image, the pedestrian in red box is inserted 10 meters in front of ego car. B-2: corresponding generated radar spectrum. Please zoom in for details.

The most important reason for this apparent contradiction is that only a few datasets provide radar data [7]. Inspired by KITTI [8] in the year 2013, most of the 3D object detection datasets include RGB camera images and LiDAR point clouds [9,10,11,12,13]. To the best of our knowledge, only nuScenes [14], Oxford Radar RobotCar [15], and Astyx HiRes2019 Datasets [16] contain radar data. Through careful analysis, we found that the radar data of the nuScenes and Astyx HiRes2019 datasets are sparse radar points instead of raw radar spectra. On the other hand, the Oxford Radar RobotCar supplies radar spectra, but without any object annotation. In short, until now, there has neither been a high-quality public dataset nor a benchmark for radar environmental perception.

Motivated by the above problems, we define automatic LiDAR-to-radar translation as the task of generating radar data from given LiDAR point clouds. It is trained with the broad set of paired LiDAR-radar samples from the Oxford RobotCar dataset, see Fig.1. A challenge that needs to be addressed is how the radar and LiDAR data are represented. Image-based representations (such as LiDAR BEV and radar spectrum images) are valid for image-to-image translation GAN frameworks that have a fixed relationship, but fewer details in raw data,

*e.g.* intensity, and height of point clouds. Otherwise, point-wise representation bases such as radar pins or point clouds are not well-suited for image-to-image translation GANs.

**Contributions** Our specific contributions are three-fold: (1) We first propose a conditional L2R GAN that can translate data from LiDAR to radar. We use an occupancy grid mask for guidance and a set of local region generators to create a more reliable link of objects between LiDAR and radar for refining small-scale regions. (2) In experiments, we demonstrate the effectiveness of our framework for the generation of raw, detailed radar spectra. Both qualitative and quantitative comparison indicate that L2R GAN outperforms previous GANs with respect to details by a large margin. (3) We show that our framework can be used for advanced data augmentation and emergency scene generation by editing the appearance of objects (such as pedestrian) in a real LiDAR scene and feeding to L2R GAN, see Fig. 1.

## 2 Background and related work

In this section, we briefly review recent existing work on data translation with conditional GAN (cGAN) and different representations of LiDAR and radar sensors.

### 2.1 Cross-Domain data translation with conditional GAN

Cross-Domain data translation, especially image-to-image translation, involves generating a new synthetic version of a given image with a specific modification, such as translating a winter landscape to summer. Generally speaking, image-to-image translation can be divided into supervised and unsupervised translation. Some early works expected to generate an output image close to a ground-truth image by reducing pixel-wise losses, for example, $L1$-loss or MSE in pixel space [17]. From 2016 on, [18] and [19] trained a conditional GAN network on paired data to translate across different image domains (like sketches to photos). In pix2pix [19], the generator is creatively designed as a U-Net architecture [20], while the discriminator classifies each $N \times N$ patch as real or fake instead of the whole image. To synthesize more photo-realistic images given an input map image, pix2pixHD uses a new multi-scale generator and discriminator [21]. In [22], authors demonstrate that conditional GAN models highly benefit from scaling up. In [23] and [24], high-resolution images are scaled using a memory bank composed of a training image segment. Spatially-adaptive normalization to transform semantic information is proposed in [25]. Very recently, LGGAN uses a local class-specific generative network with an attention fusion module to combine the multi-scaled features in the GAN [26].

Meanwhile, lots of research aims to train the network in an unsupervised way using unpaired samples from different training sets [27,28,29,30]. Furthermore,

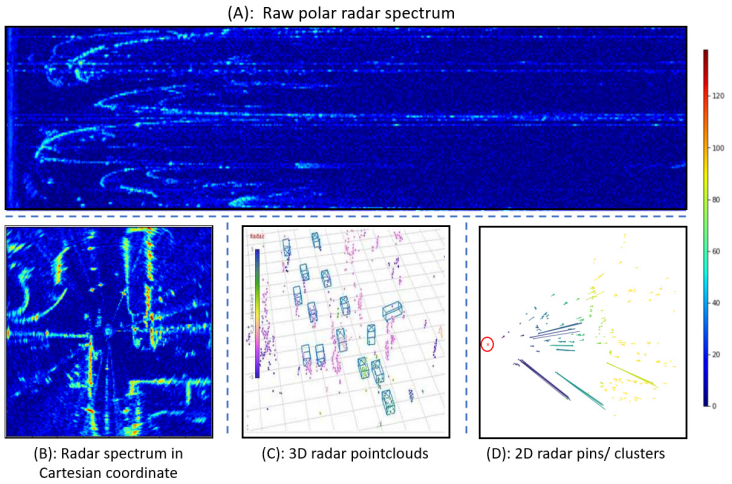


**Fig. 2.** Different Representation of radar data used in the datasets. (a): raw polar radar spectrum, Navtech CTS350-X Millimetre-Wave FMCW radar [15]; (b): the same radar spectrum in Cartesian coordinates; (c): 3D radar point clouds, Astyx 6455 HiRes radar [16]; (d): 2D radar pins/ clusters, Continental ARS40X radar [14].

[31,32,33] have presented a remarkable technique for training unsupervised image translation models via utilizing a cycle consistency loss. In 2018, to handle translation between multiple-domains without training for each pair of domains, [34] propose StarGAN, which performs this task using only a single model.

### 2.2 Representations of radar data

FMCW radars are widely used for autonomous driving with the ability to measure range (radial distance), velocity (Doppler), azimuth and received power, which is a function of the object's reflectively, size, shape, and orientation relative to the receiver, in some cases also named as radar cross section (RCS). FMCW radars continuously transmit chirp signal and receive echo signal reflected by objects. The radar measurement process is very complicated and the resulting scan is also susceptible to contamination by speckle noise, reflection, and artifacts[35]. According to the increasing levels of data abstraction and handcrafted feature extraction, radar data can be divided into the following representations: raw polar radar spectrum, radar spectrum in Cartesian coordinates, 3D radar point clouds and 2D radar pins, see Fig. 2.

The original radar raw data is in the form of a 2D array, whose row is formed by the target echo returned from each radar pulse. However, as technical secrets, such data is not available to users. Radar manufacturers use digital signal processing (DSP) algorithms such as Fast Fourier Transform (FFT) and Multiple Signal Classification (MUSIC) to obtain spectrum (range-azimuth) data under polar coordinate system [15,36]. Through coordinate transformation into a Cartesian coordinate system, we can further get the BEV spectrum images where the intensity represents the highest power reflection within a range bin.

There are several radar researches take radar spectra as input [6,4,37]. Further to ADAS applications, radar data are more heavily processed by DSP (such as clustering) and extracted to sparse 3D radar clusters or 2D radar pins [14,2,1,38].

## 3 The pix2pix and pix2pixHD baseline implementations

We propose a conditional GAN framework for generating a high-resolution radar spectrum from the 3D LiDAR point cloud which is based on the architectural ideas of pix2pix [19] and pix2pixHD [21] architectural. An illustration of the overall framework is shown in Fig. 3. In this section, as a baseline, we use the above approaches to translate the LiDAR BEV image to a radar spectrum Cartesian image, which can be formulated as a problem of image-to-image domain translation.

### 3.1 Architecture of pix2pix and pix2pixHD

The pix2pix method is a conditional GAN framework for paired image-to-image translation with an additional $L1$ loss. It consists of a U-Net [20] generator $G$ and a patch-based fully convolutional network discriminator $D$. The conditional adversarial loss with an input $x$ and ground truth $y$ is formulated as

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}\left[\log D(x, y)\right] + \mathbb{E}_{x}\left[\log(1 - D(G(x))\right]. \tag{1}$$

Moreover, training aims to find the saddle point of the objective function

$$\arg\min_{G}\max_{D}\mathcal{L}_{cGAN}(G, D) + \lambda\mathcal{L}_{pix}(G), \tag{2}$$

with a pixel-wise reconstruction loss $\mathcal{L}_{pix}$. A typical choice here is the $L_1$-norm.

The recently proposed pix2pixHD model is based on pix2pix, but has shown better results for high-resolution images synthesis. A multi-scale generator and different discriminators for multiple scales are leveraged to generate high-resolution images with details and realistic textures. The objective function is extended with the matching loss of the multiple layers' features.

We choose the same range ($80 \times 80$ meter) for both LiDAR and radar images. The reason is that the LiDAR point clouds in the far range are very sparse and few measurements are available at distances above 40m. The radar spectrum Cartesian image and BEV LiDAR image have the same representation as an image with a resolution of $N \times N = 400 \times 400$ pixels, where $N$ is the cardinality of the set of bins in the discretized range, and each pixel represents an area of $0.2 \times 0.2$ meter.

### 3.2 Drawbacks of the baseline approaches

It turns out that if we directly apply one of [19] or [21], the generated radar spectrum is quite unsatisfactory, see Fig.4. After careful analysis, we can identify four

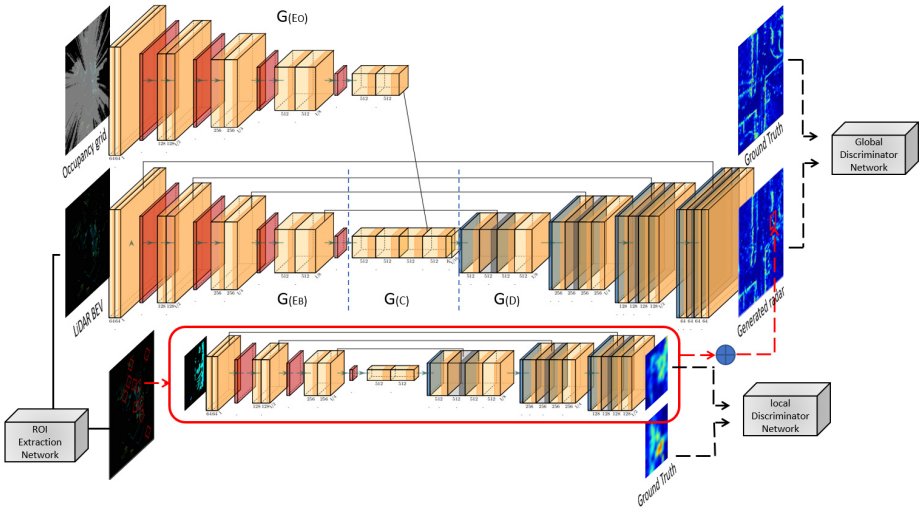


**Fig. 3.** Overview of L2R GAN network. The L2R GAN consists of four parts: a global generator, a set of local region generators, a ROI extraction network and discriminator network. The local region generator uses a U-Net framework to synthesis radar data in small-scaled ROIs , which is showed in red box. The global generator $G_{Global}$ concatenates the feature maps of occupancy grid mask and LiDAR BEV images. It also consists of four subcomponents: an occupancy grid mask encoder network $G_{(E_O)}$, a BEV image encoder network $G_{(E_B)}$, a concatenation block $G_{(C)}$, and a fusion decoder network $G_{(D)}$. See text below for details.

main reasons for this. First, due to the difference in sensor characteristics, there is no strict pixel-wise correspondence between the LiDAR BEV and Cartesian radar spectrum image. In particular cases, some objects can only be detected by either LiDAR or radar.

Second, "Black regions", such as free space and unknown regions, usually occupy most of the image area, see the radar spectrum images in Fig.2. This highly imbalanced data adds to the difficulty, which makes the GAN tend to generate more "black regions" than what would be realistic. In contrast, smaller-scaled regions (vehicles and pedestrians) can not be effectively learned by a global image-level generation, and such regions are much more critical for ADAS.

Third, since pix2pix and pix2pixHD are mainly designed for Semantic-map-guided or edge-map-guided scene generation, whose performance heavily relies on the boundaries of segments. In our case, a LiDAR BEV image alone is not able to provide boundary-like features. Furthermore, neither instance-level semantic label map nor instance maps is available.

Finally, different from a one-to-many mapping problem *e.g.* image synthesis from semantic label maps, LiDAR to radar translation is a one-to-one mapping problem. The framework should learn how to generate more realistic results instead of more diverse.

## 4 Our L2R-GAN framework

To solve the problems of the baseline approach analyzed in the previous section, we take a series of measures to increase the performance and overall quality and details of the results.

### 4.1 Occupancy-grid-mask guidance for the global generator

As discussed in the last section, "black regions", such as free space and unknown area, usually occupy most of the BEV LiDAR image area. Due to the imbalance in data distribution, the generator tends to synthesize more "black regions". Thus, to generate a more realistic radar spectrum, the corresponding region's real representation and more environment information are needed.

Inspired by [39] and [40], we assume that an occupancy grid mask of the BEV image allows the generator to better understand environmental information. The mask divides "black regions" into two classes, namely free region and unknown region. Although both regions seem similar in the BEV LiDAR image, there are obviously more radar reflections in unknown space than free space, see Fig.4. The reason behind this phenomenon is the different working principle, with phenomena such as multipath propagation, refraction, and scattering, which lets radar see part of the objects which are occluded and can not be detected by LiDAR.

To retain the basic structure of the traffic scene, we design an occupancy-grid-mask-guided global generator $G_{Global}$. The occupancy mask is generated via ray casting through the scene, which is implemented using Bresenham's line rendering algorithm [41]. For details, please see the additional material. The generator $G_{Global}$ follows an architecture in the spirit of U-Net [20] and consists of four components. The occupancy grid mask encoder network $G_{(E_O)}$ learns the features of the occupancy grid mask $M$. The BEV image encoder network $G_{(E_B)}$ is designed to encode the input BEV image $I_{BEV}$. A concatenation block $G_{(C)}$ relays the feature maps of $G_{(E_B)}$ and $G_{(E_O)}$ to the backbone framework. Finally, the fusion decoder network $G_{(D)}$ generates a coarse image of resolution $400 \times 400$. The complete layout is visualized in Fig. 3.

### 4.2 Local region generator and discriminator

To produce a truly realistic radar spectrum, a model must be able to synthesize the data of objects which occupy a smaller region, such as vehicles and pedestrians. However, most of the existing cGANs use only a global generator to capture features and texture from a large receptive field. Inspired by the idea of a coarse-to-fine generator to enhance local details [21], we separate the generator into the two sub-components $G_{Global}$ and $G_{Local}$. However, different from [21], our local generator consists of several independent local region generators, whose input is a small region of interest (ROI) instead of a whole image.

To extract ROIs from LiDAR point clouds effectively, we utilize the feature encoder network from PointPillars [42] as an extraction network. This network

is designed to convert a point cloud into a sparse pseudo-image. In our case, the feature encoder network receives the point cloud in a volume of $L \times W \times H = 80 \times 80 \times 5m$ as input and generates a pseudo-image at resolution $w \times h \times c = 400 \times 400 \times 8$ as output, where $c$ is the number of channels of the pseudo-image. We then add a 2D region proposal network (RPN) to detect ROIs in the pseudo-image. The output of whole extraction network consists of serevel ROIs, each has size of $30 \times 30 \times 3$, see Fig. 3. Notably, the Oxford Radar RobotCar dataset has no object annotation. Thus, the extraction network is trained on the nuScenes dataset, whose LiDAR sensor is the same as Oxford Radar RobotCar's.

The local region generator then processes the data on small scale ROIs extracted by the extraction net. The input of each local generation is a segment $L_i \in \mathbb{R}^{30\times30\times3}$, which is a part of BEV image that contains the segment. To control the training process and results of the local region generators, a global discriminator such as in pix2pix or multi-scale discriminator such as in pix2pixHD is insufficient. Thus, we define corresponding local region discriminators, whose input is a small-scale radar spectrum instead of a large receptive field.

For the global generator, we integrate local generator and discriminator networks which are based on the U-Net architecture, see Fig. 3. In summary, the global generator network aims to learn the large scale features of each scenario to generate globally consistent images, while the local region generator is focusing on small ROIs to enhance and refine the details in the radar spectrum. Finally, we use a fusion structure to combine the outputs of local and global generator to provide more scene details while retaining global structure. In particular, our L2R GAN is therefore capable of effectively producing high-quality radar data of each road user.

### 4.3 Objective functions

Different from other conditional GANs, the main purpose of L2R GAN is to generate a unique and as real as possible radar spectrum – no variety, but more fidelity. So the objective of L2R GAN is not only to focus on how to fool the discriminator (GAN loss), but also to reduce the difference to the corresponding ground truth. We have tried several metrics for this pixel-wise loss, such as $L1$, $L2$, and $MSE$, which we analyze in the next section.

The final objective for the global and local generators and discriminators is an expanded version of Eq. (1),

$$\arg\min_{G}\max_{D} \mathcal{L}_{cGAN}(G,D) + \lambda\mathcal{L}_{Lpix}(G) + \mu\mathcal{L}_{P}(G,D). \tag{3}$$

Here, $\mathcal{L}_P$ is a perceptual loss function known from other cGANs [21], which measures the distribution of high-level features between transformed images and ground-truth images from a discriminator. The parameters $\lambda$ and $\mu$ control the weight of pixel-wise and perceptual loss, respectively, and are different for the local and global losses. In experiments, it will turn out that the local perceptual loss does not improve results, so $\mu_{\text{local}} = 0$ for optimal results. We first train both generators separately, then jointly fine-tune them, see below for details.

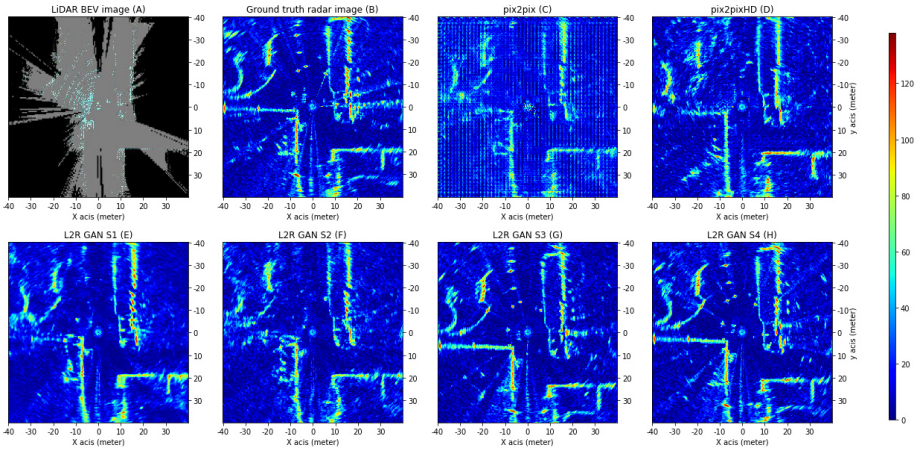


**Fig. 4.** Comparison on the Oxford robot car dataset [40]. (A) is in put LiDAR BEV image, (B) is corresponding ground truth radar image. Our method (H) generates more realistic than pix2pix (B) and pix2pixHD (C). In comparison with other baseline (E),(F), and (G), Our method (H) is closer to ground truth(B). Please zoom in for details.

## 5 Experimental results

In this section, we describe the set of experiments to evaluate our method and to demonstrate the extension of its capabilities. We then show the effectiveness of our method as a radar translator and conduct a qualitative as well as quantitative comparison against baseline methods. Due to the particularity and uniqueness of the task, we first explain the evaluation methods and metrics in Sect. 5.1. We then validate the structure of L2R GAN with a set of ablation studies to in Sect. 5.2. Finally, we show applications of our method in radar data augmentation from a novel LiDAR point clouds, and performing emergency scene generation in Sect. 6.

### 5.1 Baseline comparisons

**Implementation details.** We train the entire architecture by optimizing the objectives in Eq. (3). However, in our model, the generators $G_{Global}$ and $G_{local}$ have a considerably different number of parameters. While $G_{Global}$ is trying to learn large scale features, $G_{local}$ aims at refining the details in small scale regions. To mitigate this issue, we employ an adaptive training strategy. In order to adapt the training process at each iteration, if either discriminator's accuracy is higher than 75%, we skip its training. To avoid overfitting, we use dropout layers, which are applied to the global generator at training time. We also set different learning rates for the global discriminator, the local discriminator, the global generator, and the local generators, which are $10^{-5}$, 0.0025, $10^{-5}$, and 0.0025, respectively. We use ADAM with $\beta = 0.5$ for the optimization.

**Table 1.** qualitative experiments

| | Oxford | nuScenes |
|---|---|---|
| Ours >Pix2pix | 96.5% | 90.5% |
| Ours >Pix2pix HD | 80.5% | 85.0% |
| Ours >GT | 24.0% | no GT |

**Table 2.** quantitative experiments

| | PSNR(dB) | SSIM |
|---|---|---|
| Pix2pix | 7.722 | 0.031 |
| Pix2pix HD | 23.383 | 0.372 |
| Ours | **29.367** | **0.660** |

**Fig. 5.** Table 1 shows the results of blind randomized A/B tests on Amazon MTurk. Each entry is calculated from 200 tests made by at least 5 workers. The results indicate the percentage of comparisons in which the radar spectrum synthesized by our method are considered more realistic than the corresponding synthesized one by Pix2pix or the Pix2pixHD. Opportunity is 50%. To be noticed, for nuScenes dataset, there is no ground truth radar images. Table 2 indicates L2R GAN has less image distortion than pix2pix and pix2pixHD.

**Training, validation and test datasets.** We use the recently released Oxford Radar RobotCar Dataset [15], which consists of 280 kilometers of urban driving data under different traffic, weather and lighting conditions. The test vehicle is equipped with 2 LiDAR and 1 radar with the following specifications:

- Navtech CTS350-X Millimetre-Wave FMCW radar, 4 Hz, 400 measurements per rotation, 4.38 cm range resolution, 1.8° beamwidth, 163 m range.
- Velodyne HDL-32E 3D LIDAR, 360° HFoV, 41.3° VFoV, 32 channels, 20 Hz, 100 m range, 2 cm range resolution.

The dataset consists of several approximately 9 km trajectories in the Oxford city map. Similar to the strategy used in prior work, we manually divide the trajectories of the dataset into training, validation, and test set according to a $70 : 15 : 15$ split. So in following experiments, we use 8500 paired sample as training set, 1200 as validation and test set. Note that the LiDAR scans from each sensor are gathered at 20Hz, whereas radar streams are collected at 4Hz. Due to this temporal difference in synchronization and the dynamic environment, the translation from LiDAR to radar suffers from misalignment. We correct for this misalignment by down-sampling and interpolating the point cloud in the BEV images. In the same fashion, each radar scan is related to the closest LiDAR data in time.

For advanced data augmentation, we also use the nuScenes dataset [14] to validate the generalization ability of L2R GAN. Notably, it has no similar radar ground truth images.

**Evaluation metrics.** Evaluating the quality of synthesized radar data is an open problem and more difficult than other synthesized image. In particular, there is no common metric yet to evaluate generated radar data. To highlight the qualities of L2R GAN, we focus attention on how to generate radar data as close as possible to the ground truth. For the quantitative evaluation, we use Peak Signal to Noise Ratio (PSNR, in the range $(0, 100]$) and structural similarity (SSIM, in the range $(0, 1]$) to measure image distortion and derive the

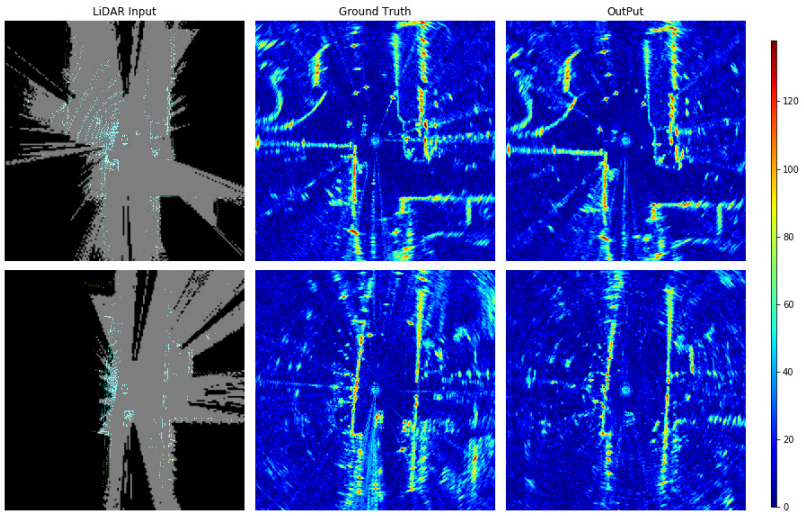


**Fig. 6.** Example radar results on the Oxford robot car dataset. Please zoom in for details.

similarity between two images [43]. The larger SSIM and PSNR, the less the image distortion.

Meanwhile, as a qualitative experiment, we also investigate a human subjective study. The evaluation metric is based on large batches of blind randomized A/B tests deployed on the Amazon Mechanical Turk platform (Amazon MTurk). To learn the characteristic of a radar spectrum, the workers were asked to first read an introduction to radar signals and browse 100 randomly selected radar spectra from the Oxford radar dataset for 10 minutes. After this, we assume that the workers have a general understanding of the characteristics and distribution of real radar data. They subsequently will be presented two images at a time, one is ground truth, the other is synthesized from the corresponding LiDAR point clouds. The workers are asked to find the real one in 8 seconds, as adopted in prior work [21].

**Baseline comparisons.** Fig. 4 and Fig. 5 report the results of baseline comparisons. Both qualitative and quantitative experiments give evidence that radar images synthesized by our approach are more accurate than the ones synthesized by Pix2pix or the Pix2pix HD. In Table 1, each entry in the table reports the percentage of comparisons in which a radar spectrum image synthesized by our approach was considered more realistic in Amazon MTurk than a corresponding one synthesized by Pix2pix or the Pix2pix HD. Fig. 6 shows more examples on the Oxford robot car dataset.

### 5.2 Ablation Analysis

**Analysis of the framework structure.** We evaluate the proposed L2R GAN in four variants S1, S2, S3, S4 as follows: (a) S1 employs only the global generator

| Method | PSNR(dB) | | | SSIM | | |
|---|---|---|---|---|---|---|
| | min | max | mean | min | max | mean |
| S1: $G_{Global}$ | 23.006 | 24.206 | 23.635 | 0.360 | 0.406 | 0.381 |
| S2: S1+$G_{Local}$ | 28.430 | 29.253 | 28.426 | 0.576 | 0.598 | 0.588 |
| S3: S1+ occupancy grid | 23.911 | 25.079 | 24.434 | 0.417 | 0.473 | 0.450 |
| S4: S3+$G_{Local}$ (/w $\mathcal{L}_1$) | **29.076** | **29.697** | **29.367** | **0.647** | 0.671 | 0.660 |
| S4 /w $\mathcal{L}_2$ | 28.261 | 29.040 | 28.781 | 0.643 | **0.674** | **0.662** |
| S4 /w $\mathcal{L}_{MSE}$ | 29.053 | 29.423 | 29.219 | 0.637 | 0.665 | 0.656 |
| S4 /w $\mathcal{L}_1 + \mathcal{L}_{P-local}$ | 27.601 | 28.211 | 27.912 | 0.543 | 0.598 | 0.572 |
| S4 /w $\mathcal{L}_2 + \mathcal{L}_{P-local}$ | 26.970 | 27.498 | 27.201 | 0.550 | 0.592 | 0.570 |
| S4 /w $\mathcal{L}_{MSE} + \mathcal{L}_{P-local}$ | 27.054 | 27.431 | 27.284 | 0.550 | 0.601 | 0.574 |

**Fig. 7.** Ablation study to evaluate different components of our framework. The upper half shows the result of different framework structure, while the lower left is analysis of loss functions. The baseline of comparison is S4 (/w $\mathcal{L}_1$), whose loss function is $\mathcal{L}_{cGAN} + \mathcal{L}_1 + \mathcal{L}_{P-Global}$.

without occupancy grid mask, (b) S2 combines the global generator without occupancy grid mask and the local region generators to produce the final results, where the local results are produced by using a point pillar based extraction network, (c) S3 uses the proposed occupancy-grid-mask-guided global generator, (d) S4 is our full model. See Fig. 7 for the evaluation result.

**Analysis of the loss functions.** Here, we show how the loss function influences the synthesis performance. For a fair comparison, we retain the same network framework and data setting as S4 and utilize a combination of different losses, see Fig. 7 for results.

Interestingly, the perceptual loss does not improve the quality of local region generators, but tends to make the training process unstable and result in collapse. We speculate that the perceptual loss may not be suitable for a small receptive field, which has few common high-level features. The experiments also show that the $L1$ loss can learn image details more effectively than $L2$ and $MSE$.

## 6 Application: data augmentation for ADAS

A big problem in ADAS is how to collect data for an emergency scenario to test a corresponding ADAS application. For example, to test Pedestrian Collision Warning (PCW), on the one hand, a sufficient number of experiments is necessary before the application is released. On the other hand, it is too dangerous to implement such a collision test under real road conditions. For this reason, researchers artificially insert real LiDAR objects into a real LiDAR scene to produce a fake dangerous traffic scenario [44]. The occluded points in the original LiDAR scene can be calculated and removed by mathematical methods, such as applying a cube map [44] or raycasting [39]. However, it is quite difficult to augment radar data in similar way. Due to refraction and scattering, the intersection of radar beams and inserted objects is much more complicated than for

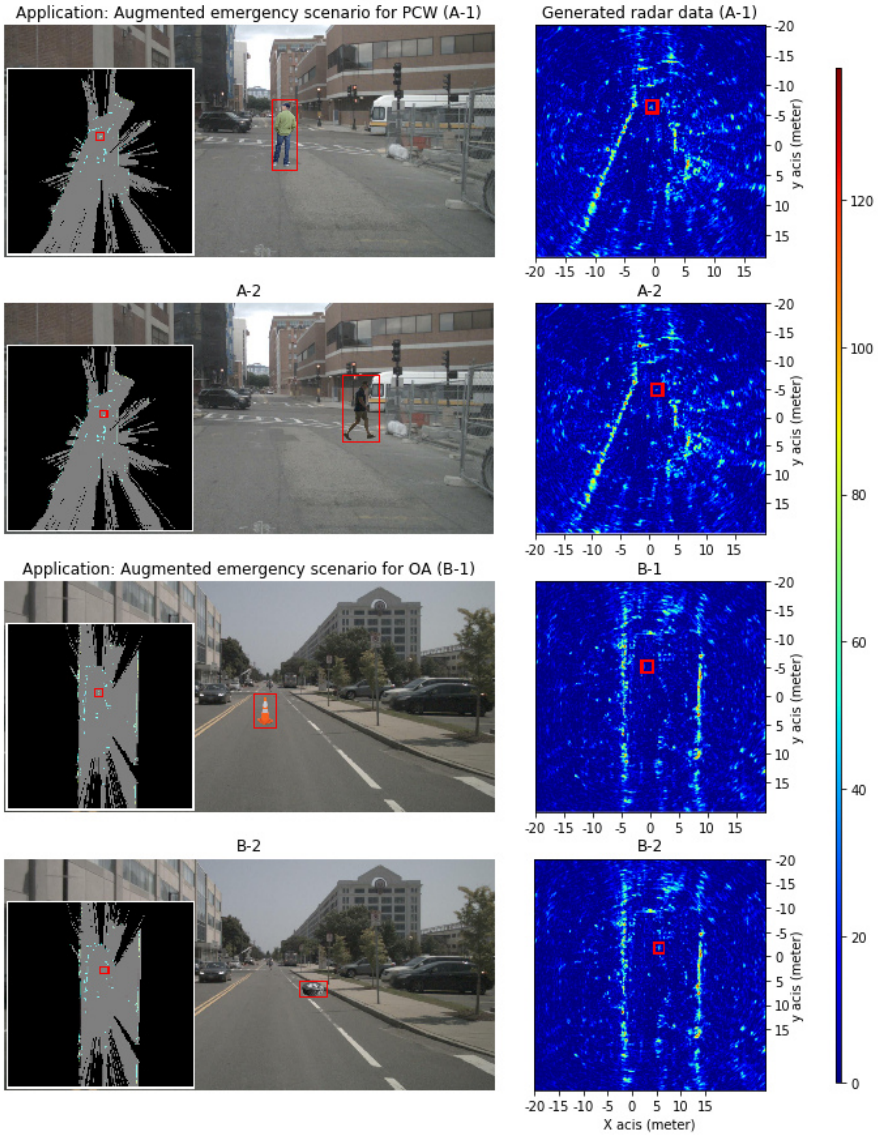


**Fig. 8.** Example radar data of augmented emergency scenario on the nuScenes dataset. In nuScenes dataset, there is no radar spectrum ground truth. A is augmented emergency scenario for Pedestrian Collision Warning (PCW): A-1 inserts a pedestrian 10 meters in front of ego car, A-2 inserts a pedestrian 2 meter east and 10 meters forward of ego car. B is augmented emergency scenario for Obstacle Avoidance (OA): B-1 inserts a traffic cone 10 meters in front of ego car, B-2 inserts a tire 2 meter east and 10 meters forward of ego car. Here camera images just help the reader understand. Please zoom in for details.

LiDAR. In the worst case, the radar wave returning from a target object can get reflected on those surfaces and result in so-called "ghost" targets that do not actually exist.

Given these observations, we propose to generate radar data of a dangerous traffic scenario by manually editing the appearance of individual objects in LiDAR data as above, then feeding the data into our L2R GAN. A GUI allows users to design their own augmented emergency scenario. To implement this idea, we collect several 3D semantically labeled objects from the nuScenes dataset (such as pedestrians, lost cargo and traffic cones) to create an object database for the user to choose from. The user can also manually select which LiDAR scenes to use as background, and where to insert a "dangerous object" of a specific class. For example, the user can add 3D points of a pedestrian 10 meters in front of the vehicle into an existing urban scenario to simulate a emergency scenario. Our L2R GAN will then automatically produce a corresponding radar spectrum. This kind of simulation data is urgently required for ADAS development and validation, which can be hardly obtained through test drive. Fig.8 shows four of these augmented scenarios.

## 7 Conclusion

In summary, we propose a new method for LiDAR-to-radar translation. Based on the pix2pix and pix2pixHD methods, our L2R GAN generates a radar spectrum image through an occupancy-grid-mask-guided global generator, a set of local region generators, a ROI extration network and discriminator networks. Results on synthetic and real radar data show promising qualitative and quantitative results which surpass the previous baseline. A L2R-GAN-based GUI also allows users to define and generate special radar data of emergency scenarios to test corresponding ADAS applications, such as pedestrian collision warning and obstacle avoidance. Our research will serve as a reference for future testing and development of various radar ADAS applications. Future investigations will focus on validating the accuracy of augmented radar data by doing experiments in the field.

## 8 Acklowdegments

This work was supported by the DFG Centre of Excellence 2117 'Centre for the Advanced Study of Collective Behaviour' (ID: 422037984).

## References


1. Lombacher, J., Hahn, M., Dickmann, J., Wöhler, C.: Potential of radar for static object classification using deep learning methods. In: 2016 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM). (2016) 1–4 1, 5
2. Schumann, O., Hahn, M., Dickmann, J., Wöhler, C.: Semantic segmentation on radar point clouds. In: 2018 21st International Conference on Information Fusion (FUSION). (2018) 2179–2186 1, 5
3. Dubé, R., Hahn, M., Schütz, M., Dickmann, J., Gingras, D.: Detection of parked vehicles from a radar based occupancy grid. In: IEEE Intelligent Vehicles Symposium (IV). (2014) 1415–1420 1
4. Cen, S.H., Newman, P.: Radar-only ego-motion estimation in difficult settings via graph matching. In: 2019 International Conference on Robotics and Automation (ICRA). (2019) 298–304 1, 5
5. Bartsch, A., Fitzek, F., Rasshofer, R.: Pedestrian recognition using automotive radar sensors. Advances in Radio Science: ARS **10** (2012) 1
6. Dong, X., Wang, P., Zhang, P., Liu, L.: Probabilistic orientated object detection in automotive radar. arXiv preprint arXiv:2004.05310 (2020) 1, 5
7. Feng, D., Haase-Schütz, C., Rosenbaum, L., Hertlein, H., Glaeser, C., Timm, F., Wiesbeck, W., Dietmayer, K.: Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. IEEE Transactions on Intelligent Transportation Systems (2020) 2
8. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: The kitti vision benchmark suite. URL http://www. cvlibs. net/datasets/kitti (2015) 2
9. Huang, X., Cheng, X., Geng, Q., Cao, B., Zhou, D., Wang, P., Lin, Y., Yang, R.: The apolloscape dataset for autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. (2018) 954–960 2
10. Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., et al.: Scalability in perception for autonomous driving: Waymo open dataset. In: cvpr. (2020) 2446–2454 2
11. Pham, Q.H., Sevestre, P., Pahwa, R.S., Zhan, H., Pang, C.H., Chen, Y., Mustafa, A., Chandrasekhar, V., Lin, J.: A* 3d dataset: Towards autonomous driving in challenging environments. arXiv preprint arXiv:1909.07541 (2019) 2
12. Hwang, S., Park, J., Kim, N., Choi, Y., So Kweon, I.: Multispectral pedestrian detection: Benchmark dataset and baseline. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2015) 1037–1045 2
13. Jung, H., Oto, Y., Mozos, O.M., Iwashita, Y., Kurazume, R.: Multi-modal panoramic 3d outdoor datasets for place categorization. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). (2016) 4545–4550 2
14. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuscenes: A multimodal dataset for autonomous driving. arXiv preprint arXiv:1903.11027 (2019) 2, 4, 5, 10
15. Barnes, D., Gadd, M., Murcutt, P., Newman, P., Posner, I.: The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset. arXiv preprint arXiv:1909.01300 (2019) 2, 4, 10
16. Meyer, M., Kuschk, G.: Automotive radar dataset for deep learning based 3d object detection. In: 2019 16th European Radar Conference (EuRAD). (2019) 129–132 2, 4

17. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence **38** (2015) 295–307 3
18. Sangkloy, P., Lu, J., Fang, C., Yu, F., Hays, J.: Scribbler: Controlling deep image synthesis with sketch and color. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017) 5400–5409 3
19. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017) 1125–1134 3, 5
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. (2015) 234–241 3, 5, 7
21. Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018) 8798–8807 3, 5, 7, 8, 11
22. Brock, A., Donahue, J., Simonyan, K.: Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096 (2018) 3
23. Qi, X., Chen, Q., Jia, J., Koltun, V.: Semi-parametric image synthesis. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018) 8808–8816 3
24. Chen, Q., Koltun, V.: Photographic image synthesis with cascaded refinement networks. In: IEEE International Conference on Computer Vision (ICCV). (2017) 1511–1520 3
25. Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y.: Semantic image synthesis with spatially-adaptive normalization. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2019) 2337–2346 3
26. Tang, H., Xu, D., Yan, Y., Torr, P.H., Sebe, N.: Local class-specific and global image-level generative adversarial networks for semantic-guided scene generation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2020) 7870–7879 3
27. Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. In: Advances in neural information processing systems. (2017) 700–708 3
28. Liu, M.Y., Tuzel, O.: Coupled generative adversarial networks. In: Advances in neural information processing systems. (2016) 469–477 3
29. Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., Webb, R.: Learning from simulated and unsupervised images through adversarial training. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017) 2107–2116 3
30. Bousmalis, K., Silberman, N., Dohan, D., Erhan, D., Krishnan, D.: Unsupervised pixel-level domain adaptation with generative adversarial networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017) 3722–3731 3
31. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE International Conference on Computer Vision (ICCV). (2017) 2223–2232 4
32. Yi, Z., Zhang, H., Tan, P., Gong, M.: Dualgan: Unsupervised dual learning for image-to-image translation. In: IEEE International Conference on Computer Vision (ICCV). (2017) 2849–2857 4