

Range-Doppler map upsampling for single channel chirp sequence radar using Deep Learning

Master thesis



Zheming Yin

Institute of Signal Processing and
System Theory

University of Stuttgart

24.04.2025

- Single channel chirp sequence radar
- Range-Doppler map upsampling

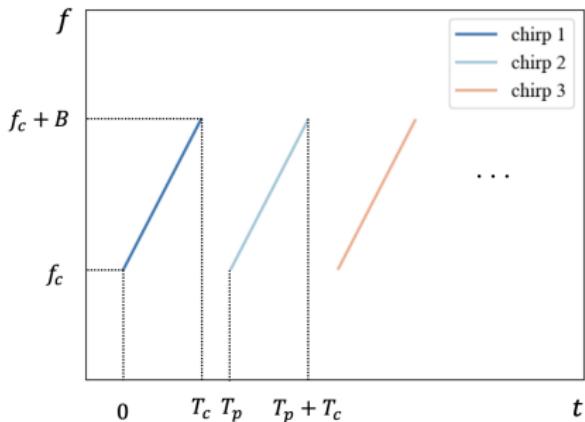


Figure 0.1. The principle of the chirp sequence radar based on the frequency modulated continuous wave (FMCW) radar

- Importance in the indoor localization
- Limited resolution by the radar system, cost, regulation, etc.
- Specified for the range-Doppler map
- Wide dynamic range of the amplitude

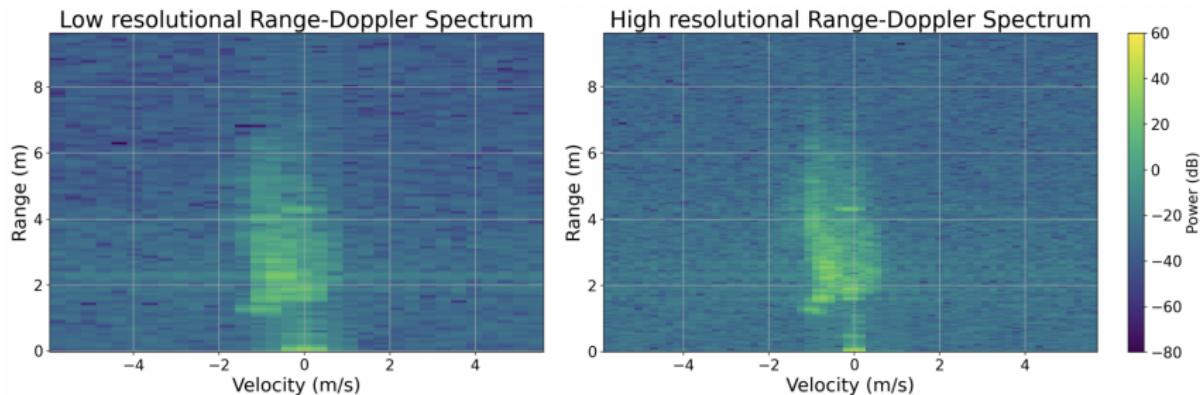


Figure 0.2. An example of the paired low-resolution and high-resolution range-Doppler maps.

- Importance in the indoor localization
- Limited resolution by the radar system, cost, regulation, etc.
- Specified for the range-Doppler map
- Wide dynamic range of the amplitude

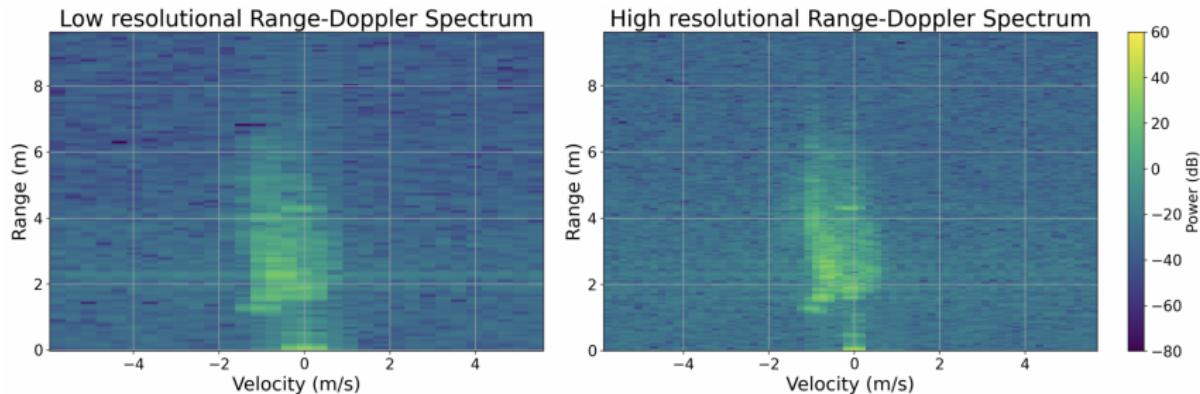


Figure 0.2. An example of the paired low-resolution and high-resolution range-Doppler maps.

Table of Contents



Dataset

Dataset collection

Dataset loading

Dataset processing

Models

Dimension processing layer

Upsampling layer

Multiple models

Loss functions

Evaluation

Summary & Outlook

Table of Contents



Dataset

Dataset collection

Dataset loading

Dataset processing

Models

Dimension processing layer

Upsampling layer

Multiple models

Loss functions

Evaluation

Summary & Outlook

Hardware & Initialization

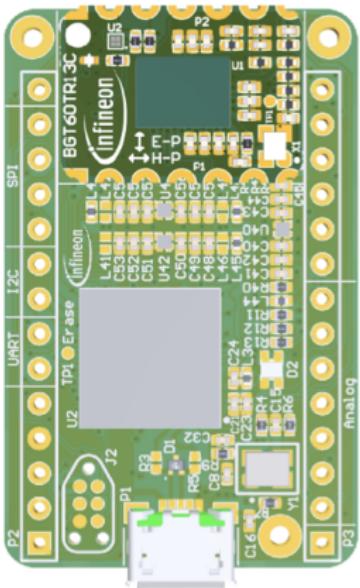
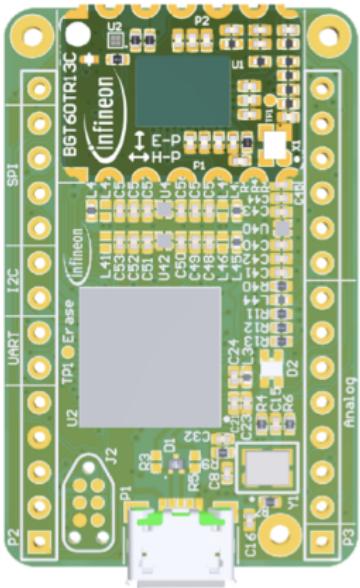


Figure 1.1. BGT60TR13C radar
from Infineon [1]

- #Chirps $N_c = 64$
- #Samples of chirp $N_s = 512$
- Shape (512, 64)
- Maximum range $r_{max} = 9.6 \text{ m}$
- Range resolution $\Delta r = 0.0375 \text{ m}$
- Maximum velocity $v_{max} = 5.68 \text{ m/s}$
- Velocity resolution $\Delta v = 0.178 \text{ m/s}$

Hardware & Initialization



- #Chirps $N_c = 64$
- #Samples of chirp $N_s = 512$
- Shape (512, 64)
- Maximum range $r_{max} = 9.6 \text{ m}$
- Range resolution $\Delta r = 0.0375 \text{ m}$
- Maximum velocity $v_{max} = 5.68 \text{ m/s}$
- Velocity resolution $\Delta v = 0.178 \text{ m/s}$

Figure 1.1. BGT60TR13C radar
from Infineon [1]

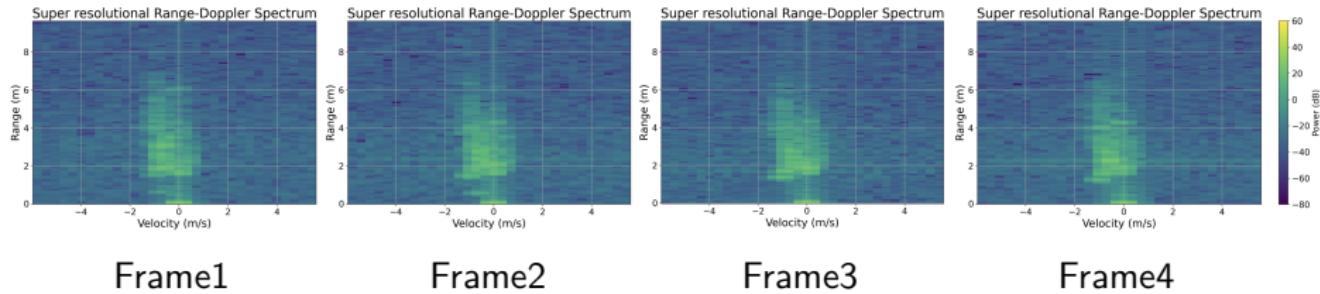


Figure 1.2. Consecutive frames during the collection

- Corridor environment
- Over 90,000 frames
- Size 16.4 GB
- Dynamic motion
- Different tempos
- Different directions

Resampling

- Range resolution

$$\Delta r = \frac{c}{2B} \quad (1)$$

- Velocity resolution

$$\Delta v = \frac{c}{2f_c} \cdot \frac{1}{T_p \cdot N_c} \quad (2)$$

- Maximum range

$$r_{\max} = \frac{c \cdot F_s}{2S} \quad (3)$$

- Maximum velocity

$$v_{\max} = \frac{\lambda}{4 \cdot T_p} \quad (4)$$

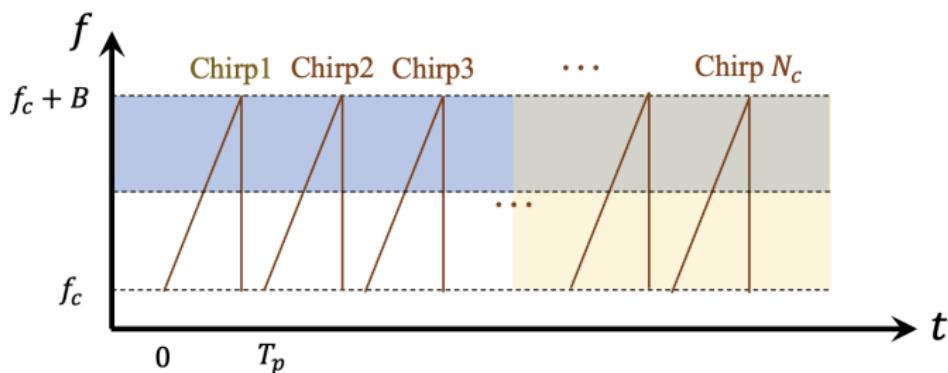


Figure 1.3. Downsampling by factor 2

Range-Doppler processing (RDP)

- Real-valued fast Fourier transform (rFFT) on range axis
 - Keep the positive frequency part and an offset
- Fast Fourier transform (FFT) on velocity axis
- High-resolution shape from (512, 64) to (257, 64)
- Low-resolution shape from (256, 32) to (129, 32)

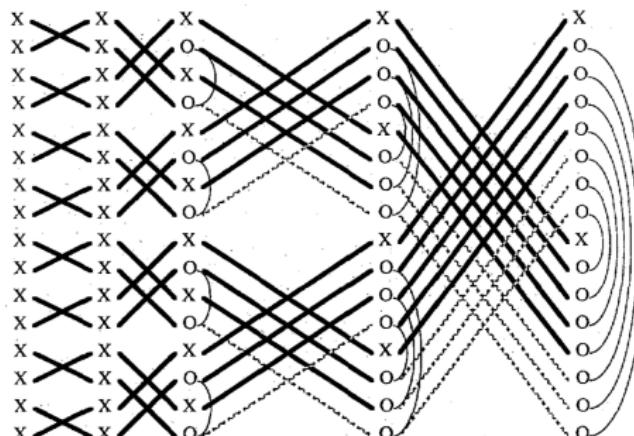


Figure 1.4. Overview of the FFT in the case of real-valued inputs [2]

- Input data representation
 - Real and imaginary representation
 - Amplitude representation
 - Amplitude and phase representation
- Logarithm
- Normalization
 - Normalization on the amplitude

$$\text{normalized data} = \frac{\text{data}}{\max} \quad (5)$$

$$\text{normalized data} = \frac{\text{data} - \min}{\max - \min} \quad (6)$$

- Normalization on the phase

Table of Contents



Dataset

Dataset collection

Dataset loading

Dataset processing

Models

Dimension processing layer

Upsampling layer

Multiple models

Loss functions

Evaluation

Summary & Outlook

- No dimension processing layer
 - Low-resolution shape keeps (129, 32)
 - After upsampling, shape from (129 , 32) to (258, 64)
 - Discard last sample, from (258, 64) to (257, 64)
- Padding along the range axis
 - Low-resolution shape from (129, 32) to (130, 32)
 - After upsampling, shape from (130 , 32) to (260, 64)
 - Discard last three samples, from (260, 64) to (257, 64)
- Convolutional layer with kernel size (2, 1) and stride 1
 - Low-resolution shape from (129, 32) to (128, 32)
 - After upsampling, shape from (128 , 32) to (256, 64)
 - Transposed convolutional layer from (256, 64) to (257, 64)

Upsampling layer

- Transposed convolutional layer
- Pixel shuffle layer (depth-to-space method)
 - Shape from $(N, C \times r^2, H, W)$ to $(N, C, H \times r, W \times r)$

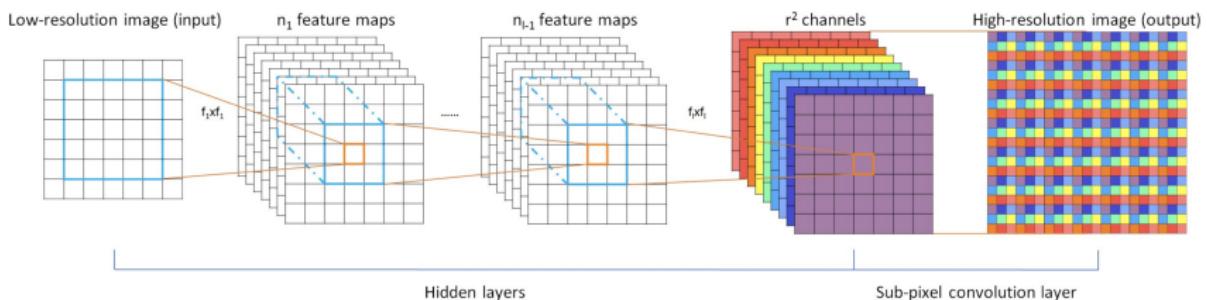


Figure 2.1. Pixel shuffle layer [3]

Interpolation model

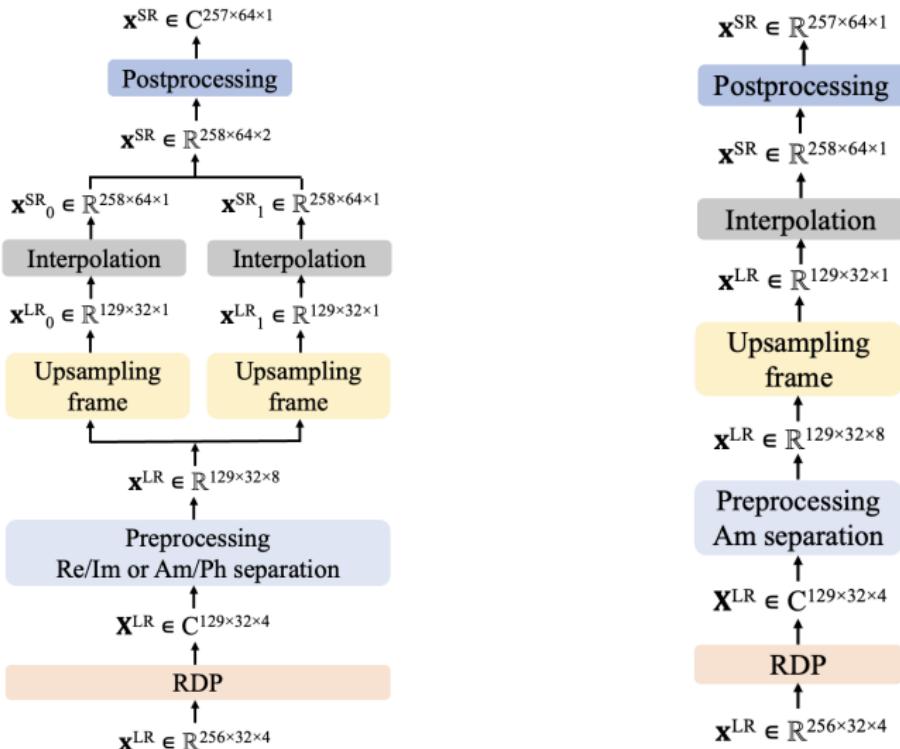


Figure 2.2. Interpolation model according to the input data representation

- Convolutional neural network (CNN) model

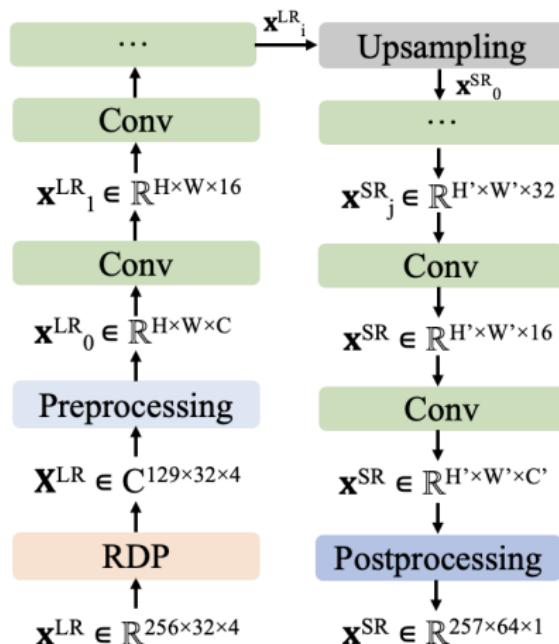


Figure 2.3. CNN model

UNet model

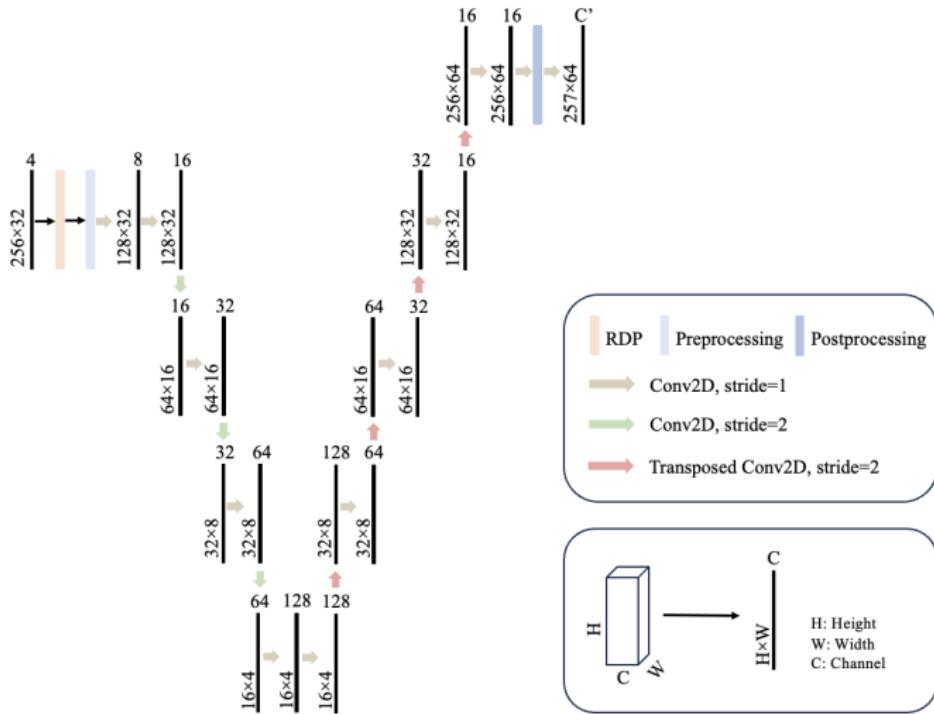


Figure 2.4. UNet model

UNet concat model [4]

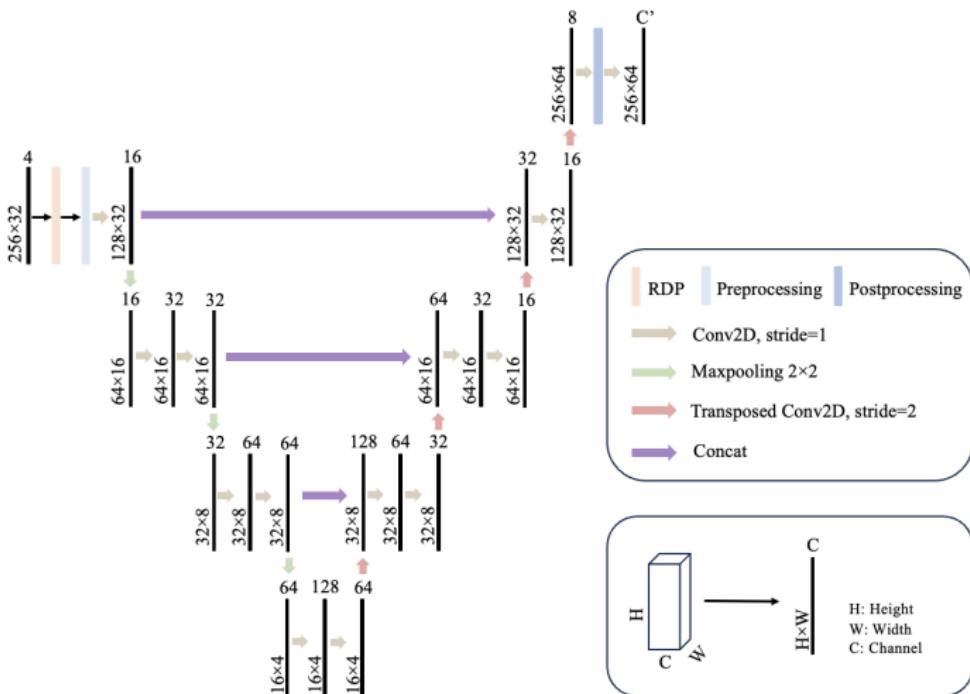


Figure 2.5. UNet concat model, adapted from [4]

DP-TF Transformer model

- Dual-Path Time-Frequency (DP-TF) Transformer model [5]

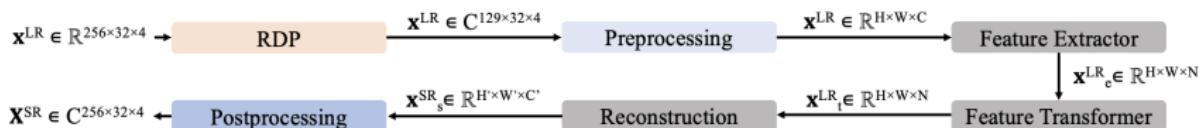


Figure 2.6. DP-TF Transformer model, adapted from [5]

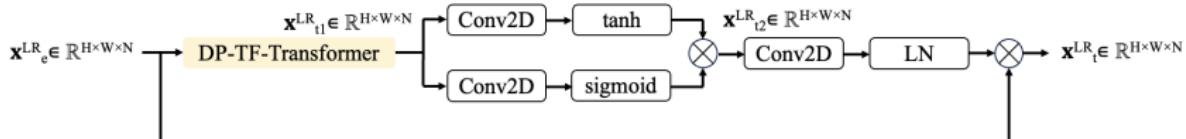


Figure 2.7. Feature Transformer block, adapted from [5]

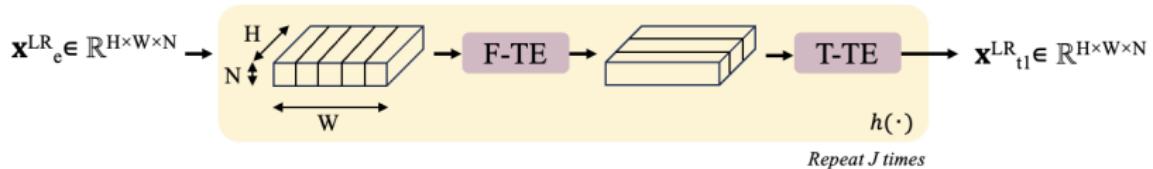


Figure 2.8. DP-TF-Transformer block, adapted from [5]

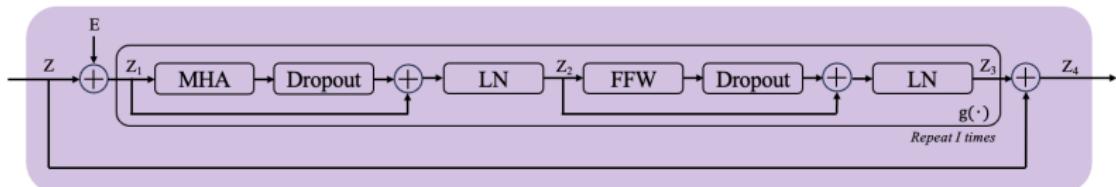


Figure 2.9. Transformer encoder (TE) block, adapted from [5]

SwinIR Transformer architecture

- Image restoration with Swin Transformer (SwinIR) architecture [6]

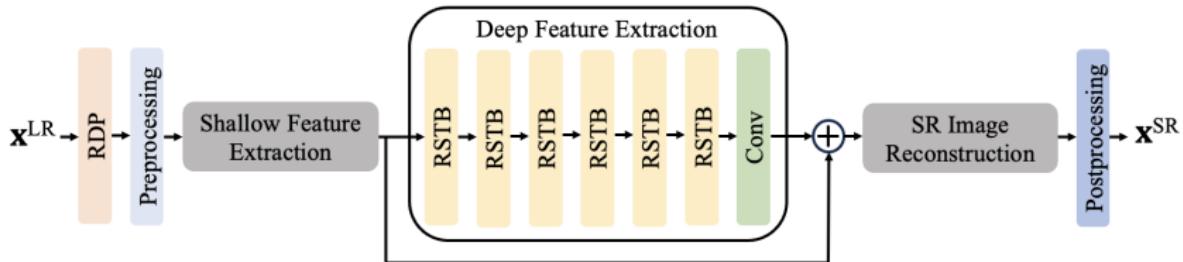


Figure 2.10. SwinIR Transformer architecture, adapted from [6]

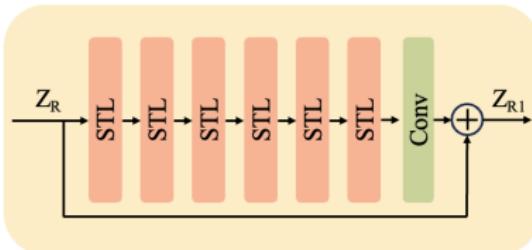


Figure 2.11. Residual Swin Transformer blocks (RSTB), adapted from [6]

SwinIR Transformer models

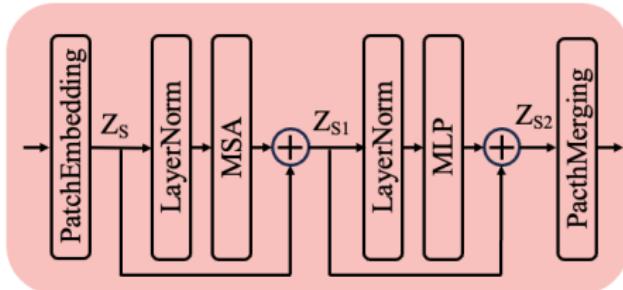


Figure 2.12. SwinIR+Swin: Swin Transformer layers (STL) block, adapted from [6]

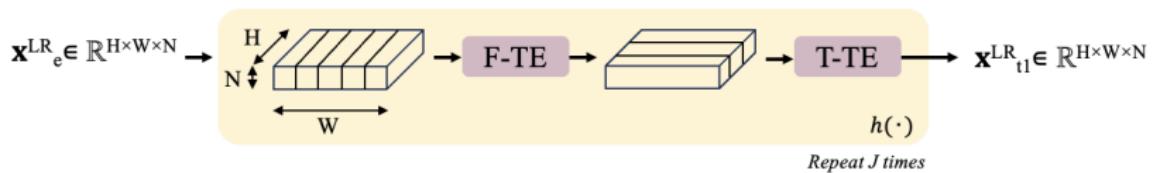


Figure 2.13. SwinIR+DP: DP-TF Transformer block, adapted from [5]

- Conditional generative adversarial network (cGAN) model [7]
- Generator
- Discriminator

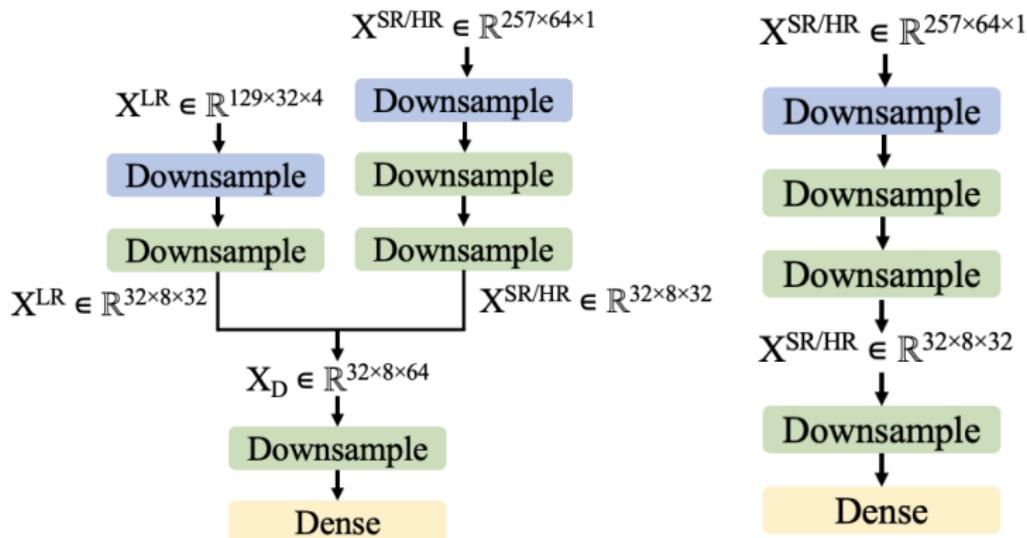


Figure 2.14. Discriminator models with and without low-resolution inputs

Table of Contents



Dataset

Dataset collection

Dataset loading

Dataset processing

Models

Dimension processing layer

Upsampling layer

Multiple models

Loss functions

Evaluation

Summary & Outlook

- Mean squared error (MSE)

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (7)$$

$$\mathcal{L}_{\text{MSE}} = \mathcal{L}_{\text{MSE, Re/Amp}} + \lambda \times \mathcal{L}_{\text{MSE, Im/Ph}} \quad (8)$$

- Weighted MSE (WMSE)

$$\mathcal{L}_{\text{WMSE, Amp}} = \frac{1}{n} \sum_{i=1}^n \frac{(A_i - \hat{A}_i)^2}{A_i} \quad (9)$$

$$\mathcal{L}_{\text{WMSE}} = \mathcal{L}_{\text{WMSE, Amp}} + \lambda \times \mathcal{L}_{\text{MSE, Ph}} \quad (10)$$

- Signal-to-distortion ratio (SDR) [8]

$$\mathcal{L}_{\text{SDR}} = 10 \log_{10} \left(\frac{\|A\|^2}{\|A - \hat{A}\|^2} \right) \quad (11)$$

- Logarithmic spectral distance (LSD) [9]

$$\mathcal{L}_{\text{LSD, Amp}} = \left\{ \frac{1}{N} \sum_{n=1}^N \left[\log_{10} A(n) - \log_{10} \hat{A}(n) \right]^2 \right\}^{\frac{1}{2}} \quad (12)$$

$$\mathcal{L}_{\text{LSD}} = \mathcal{L}_{\text{LSD, Amp}} + \lambda \times \mathcal{L}_{\text{MSE, Ph}} \quad (13)$$

Mask in LSD loss

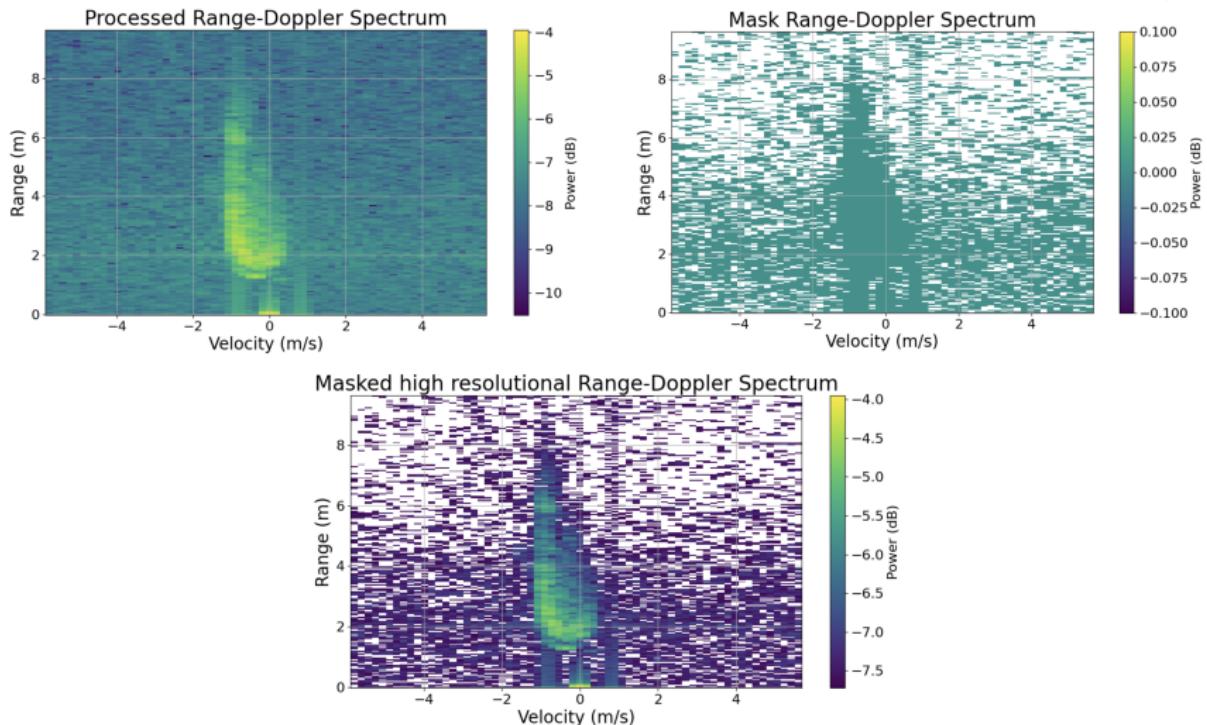


Figure 3.1. Mask in LSD loss function. From left to right: processed high-resolution data, mask; below: masked data.

- Phase-aware logarithmic spectral distance (PLSD) [9]

$$\mathcal{L}_{\text{PLSD}} = \left\langle \left| \log_{10} \left| \frac{\hat{Y}}{Y} \right| \right| \times (2 - \cos(\varphi_{\hat{Y}} - \varphi_Y)) \right\rangle \quad (14)$$

- VGG perceptual loss [10]

$$\mathcal{L}_{\text{Perceptual}}(\hat{Y}, Y) = \frac{1}{HWC} \left\| \phi(\hat{Y}) - \phi(Y) \right\|_2^2 \quad (15)$$

- Loss combination, such as

$$\mathcal{L}_{\text{Combination}} = \mathcal{L}_{\text{LSD}} + \lambda_c \times \mathcal{L}_{\text{Perceptual}} \quad (16)$$

- Adversarial loss

 - Generator loss

$$\mathcal{L}_{\text{Gen}} = \mathcal{L}_{\text{Combination}} + \lambda_g \times \|1 - P_{\text{Super}}\|_2 \quad (17)$$

 - Discriminator loss

$$\mathcal{L}_{\text{Disc}} = \|1 - P_{\text{High}}\|_2 + \|0 - P_{\text{Super}}\|_2 \quad (18)$$

Table of Contents



Dataset

Dataset collection

Dataset loading

Dataset processing

Models

Dimension processing layer

Upsampling layer

Multiple models

Loss functions

Evaluation

Summary & Outlook

- Evaluation one by one
- Only amplitude loss in the evaluation of loss functions
- Convert back the logarithm and normalization
- Small model and data subset
- Criteria: Lower evaluation loss and better visual effect

Table 4.1: Evaluation losses of different models in the case of the MSE and common processing methods.

Models	#Params	MSE	SDR	LSD	WMSE	Perceptual
Interpolation	0	3.570	-3.620	0.750	0.774	28.010
CNN	103,684	2.164	-5.028	0.441	3.599	25.133
UNet	93,162	1.352	-2.872	0.600	1.034	22.014
UNet concat	96,746	1.275	-3.609	0.552	0.968	21.847
DP	113,964	3.212	-4.926	0.611	0.163	23.583
SwinIR+DP	97,252	1.754	-5.787	0.503	0.314	22.764
SwinIR+Swin	111,624	1.541	-4.505	0.546	1.104	26.381

Models evaluation

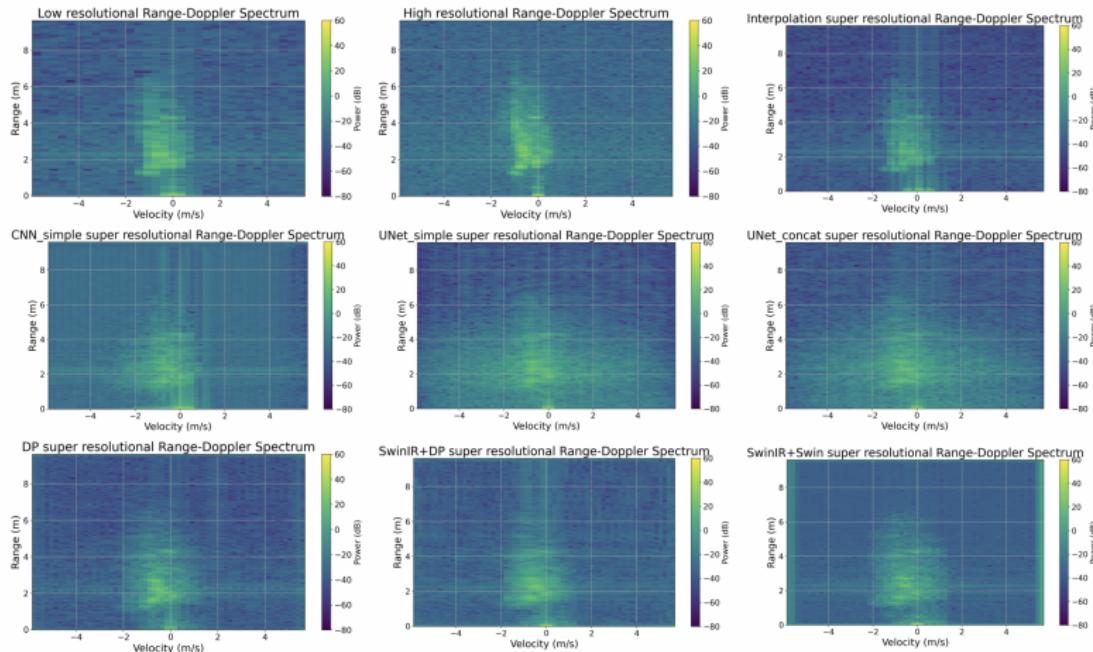


Figure 4.1. Range-Doppler maps in the case of MSE and common processing methods, from left to right, first row: low-resolution, high-resolution, interpolation; the second row: CNN, UNet, UNet concat; third row: DP, SwinIR+DP, SwinIR+Swin.

Processing methods evaluation

Table 4.2: Evaluation losses of different input data representations

Representation types	MSE	SDR	LSD	WMSE	Perceptual
Real/Imaginary	3.435	-4.958	0.619	0.155	24.682
Amplitude	3.136	-7.778	0.385	0.181	18.101
Amplitude/Phase	3.213	-5.068	0.493	0.241	20.126

Table 4.3: Evaluation losses of different dimension processing types

Dimension	MSE	SDR	LSD	WMSE	Perceptual
No processing	3.100	-6.312	0.427	0.194	19.081
Padding	3.113	-6.611	0.420	0.190	19.092
Convolution	2.349	-5.908	0.451	0.206	16.259

Table 4.4: Evaluation losses of different upsampling types

Upsampling	#Params	MSE	SDR	LSD	WMSE	Perceptual
Transposed	113,866	3.113	-6.611	0.420	0.190	19.092
Pixel shuffle	48,042	2.560	-4.206	0.506	0.269	17.918

Table 4.5: Evaluation losses of the effect of the logarithm types

Logarithm types	MSE	SDR	LSD	WMSE	Perceptual
No logarithm	2.560	-4.206	0.506	0.269	17.918
With logarithm	1.740	-9.251	0.295	0.130	17.636

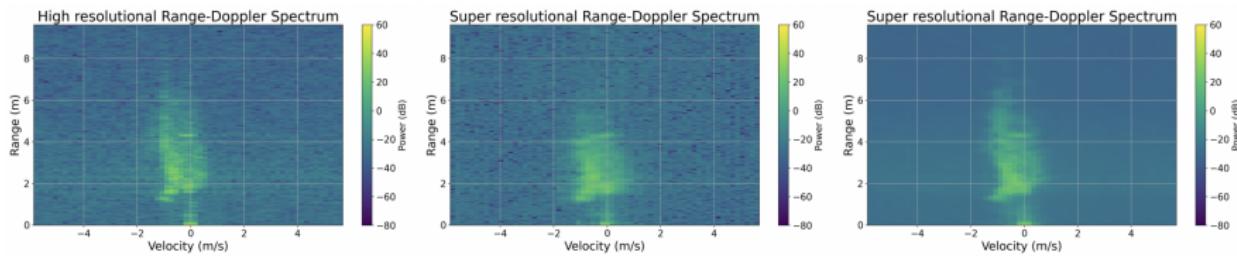


Figure 4.2. From left to right, high-resolution range-Doppler map, super-resolution range-Doppler maps of the cases without and with logarithm operation, respectively.

Table 4.6: Evaluation losses of different amplitude normalization types

Amplitude normalization	MSE	SDR	LSD	WMSE	Perceptual
No normalization	1.740	-9.251	0.295	0.130	17.636
Normalization in (-1, 1)	1.711	-6.841	0.349	0.108	14.770
Normalization in (0, 1)	1.073	-6.327	0.358	0.119	14.639

Table 4.7: Evaluation losses of the effect of the angle normalization type

Angle normalization	MSE	SDR	LSD	WMSE	Perceptual
No normalization	1.073	-6.327	0.358	0.119	14.639
With normalization	1.330	-6.068	0.363	0.117	15.204

Table 4.8: Evaluation losses of different training loss functions, horizontal axis is the evaluation loss functions, vertical axis is the training loss functions.

Loss evaluation	MSE	SDR	LSD	WMSE	Perceptual
MSE	1.073	-6.327	0.358	0.119	14.639
SDR	1.110	-5.683	0.370	0.121	14.163
LSD	1.232	-6.736	0.351	0.108	14.198
PLSD	1.424	-6.434	0.356	0.124	16.999
WMSE	1.279	-5.967	0.365	0.131	17.478
Perceptual	6.174	-4.129	0.417	0.393	13.882

Training loss functions evaluation

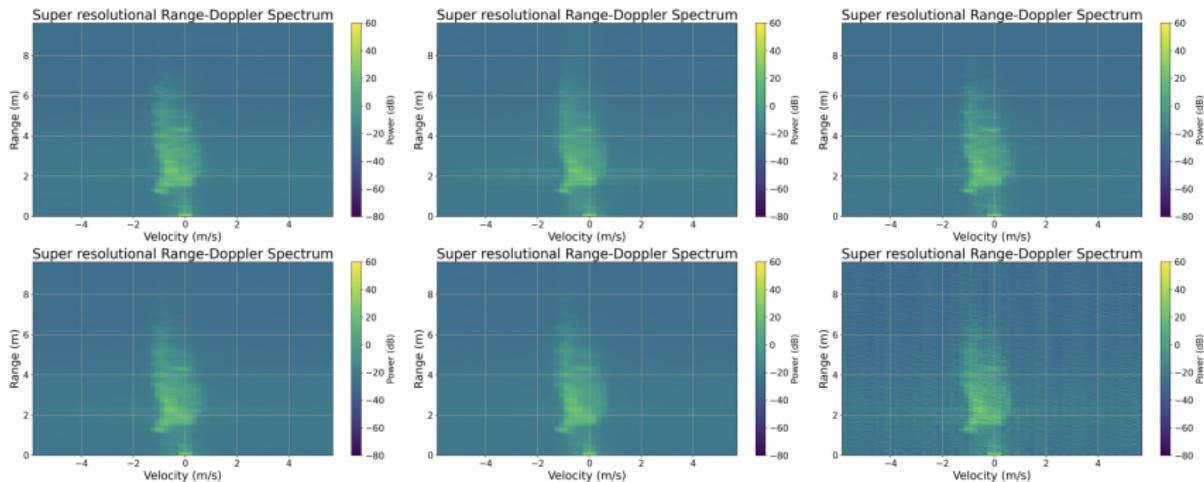


Figure 4.3. Super-resolution range-Doppler maps of different training loss functions, from left to right, in the first row MSE, SDR and LSD, in the second row PLSD, WMSE and perceptual loss functions.

$$\mathcal{L}_{\text{Combination}} = \mathcal{L}_{\text{LSD}} + \lambda_c \times \mathcal{L}_{\text{Perceptual}} = \mathcal{L}_{\text{LSD}} + 0.5 \times \mathcal{L}_{\text{Perceptual}} \quad (19)$$

$$\mathcal{L}_{\text{Gen}} = \mathcal{L}_{\text{LSD}} + 0.5 \times \mathcal{L}_{\text{Perceptual}} + \lambda_g \times \mathcal{L}_{\text{GAN, Gen}} \quad (20)$$

Table 4.9: Evaluation losses of cGAN model with and without low-resolution range-Doppler map compared with only generator, the #Params of generator is 48,042 while that of discriminator is 76,267.

cGAN	MSE	SDR	LSD	WMSE	Perceptual
With low-resolution input in discriminator					
$\lambda_g=1e-2$	1.339	-5.103	0.385	0.177	11.926
Without low-resolution input in discriminator					
$\lambda_g=1e-2$	1.640	-5.690	0.374	0.169	11.881
Only generator					
DP	1.731	-5.375	0.380	0.179	12.320

Table 4.10: Evaluation losses of different numbers of frames.

#Frames evaluation	MSE	SDR	LSD	WMSE	Perceptual
#Frames = 1					
DP	1.379	-5.239	0.381	0.191	12.778
cGAN	1.950	-3.930	0.409	0.236	13.063
#Frames = 4					
DP	1.731	-5.375	0.380	0.179	12.320
cGAN	1.640	-5.690	0.374	0.169	11.881

Hyperparameter tuning

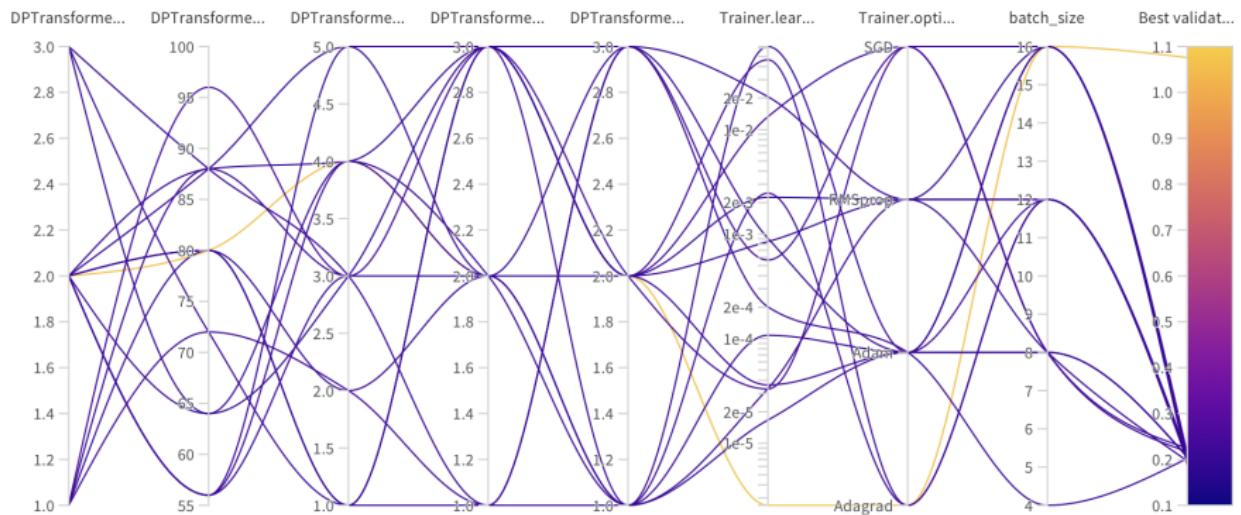


Figure 4.4. Hyperparameter tuning by WandB [11]

Final result

Table 4.11: Evaluation losses with large models and whole dataset, where the #Params of generator is 329,410 and that of discriminator is 766,187.

Final result	MSE	SDR	LSD	WMSE	Perceptual
DP	0.968	-6.008	0.363	0.123	9.079
cGAN	0.803	-5.619	0.371	0.120	8.740

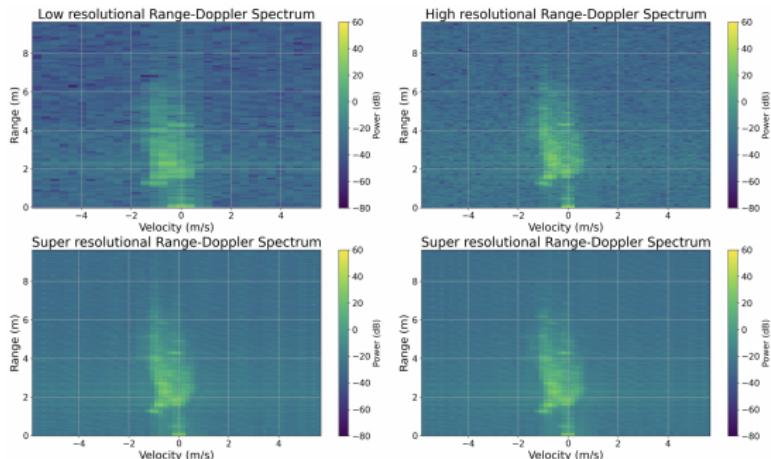


Figure 4.5. Super-resolution range-Doppler maps trained by large DP-TF Transformer and cGAN models with the whole dataset.

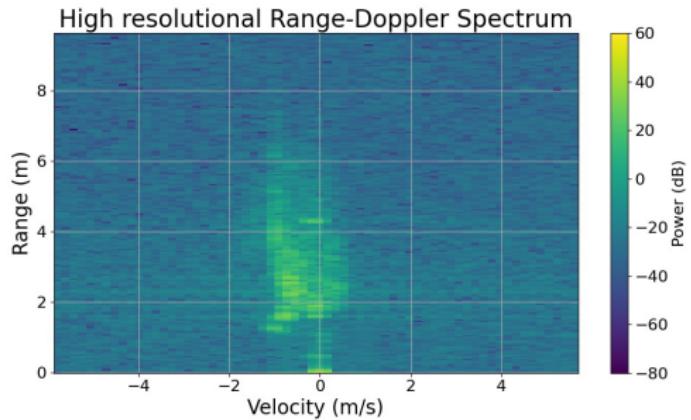


Figure 4.6. Visualization of range-Doppler map in the early epochs

Table of Contents



Dataset

Dataset collection

Dataset loading

Dataset processing

Models

Dimension processing layer

Upsampling layer

Multiple models

Loss functions

Evaluation

Summary & Outlook

- Dataset collection
- DP-TF Transformer model with tuned hyperparameters
- Processing methods
 - Amplitude and phase representation
 - Padding to deal with the range axis
 - Pixel shuffle layer to upsample
 - Logarithm and normalization on amplitude
- Loss combination with LSD and perceptual loss
$$\mathcal{L}_{\text{Combination}} = \mathcal{L}_{\text{LSD}} + 0.5 \times \mathcal{L}_{\text{Perceptual}}$$
- cGAN, where the generator is DP-TF Transformer model
$$\mathcal{L}_{\text{Gen}} = \mathcal{L}_{\text{LSD}} + 0.5 \times \mathcal{L}_{\text{Perceptual}} + 0.01 \times \|1 - P_{\text{Super}}\|_2$$

- Enrich the dataset further, such as more environments
- More permutations and evaluations on the combination
- Model optimization, such as division into different shapes
- Other loss functions

- [1] Infineon Technologies AG. *BGT60TR13C / XENSIV 60GHz radar sensor for advanced sensing - infineon technologies*. URL: <https://www.infineon.com/cms/en/product/sensor/radar-sensors/radar-sensors-for-iot/60ghz-radar/bgt60tr13c/>.
- [2] H. Sorensen u. a. "Real-valued fast fourier transform algorithms". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 35 (1987), S. 849–863. URL: <http://ieeexplore.ieee.org/document/1165220/>.
- [3] Wenzhe Shi u. a. *Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network*. 2016. URL: <http://arxiv.org/abs/1609.05158>.
- [4] Akash Prabhakara u. a. "High resolution point clouds from mmWave radar". In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, S. 4135–4142. URL: <https://ieeexplore.ieee.org/document/10161429/>.
- [5] Sven Hinderer. "Blind source separation of radar signals in time domain using deep learning". In: *2022 23rd International Radar Symposium (IRS)*. IEEE, 2022, S. 486–491. URL: <https://ieeexplore.ieee.org/document/9904990/>.
- [6] Jingyun Liang u. a. *SwinIR: Image restoration using swin transformer*. 2021. URL: <http://arxiv.org/abs/2108.10257>.
- [7] Keras Team. *Keras documentation: Conditional GAN*. URL: https://keras.io/examples/generative/conditional_gan/.
- [8] Jonathan Le Roux u. a. *SDR - half-baked or well done?* 2018. URL: <http://arxiv.org/abs/1811.02508>.
- [9] Sebastian Braun und Ivan Tashev. *A consolidated view of loss functions for supervised deep learning-based speech enhancement*. 2020. URL: <http://arxiv.org/abs/2009.12286>.
- [10] Justin Johnson, Alexandre Alahi und Li Fei-Fei. *Perceptual losses for real-time style transfer and super-resolution*. 2016. URL: <http://arxiv.org/abs/1603.08155>.
- [11] Weights & biases. 2025. URL: <httpss://wandb.ai/site>.

Thank you for your attention