Technology Innovation Institute

Transformer-Based Multi-Modal Deep Learning for Sensing-Aid Beam Prediction

ITU AI/ML in 5G Grand Challenge 2022

Qiyang Zhao, Yu Tian, Zine Kherroubi, Fouzi Boukhalfa*

Thanks to Kebin Wu

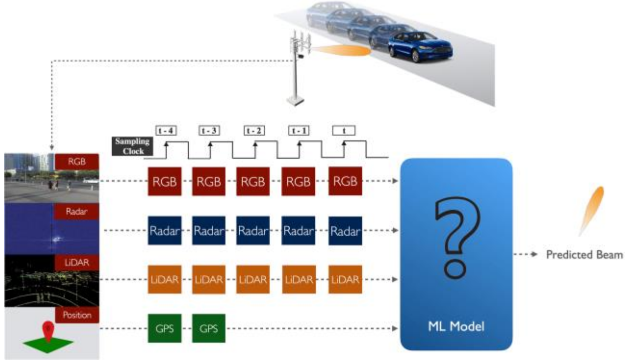
*Team members contributed equally to this work.

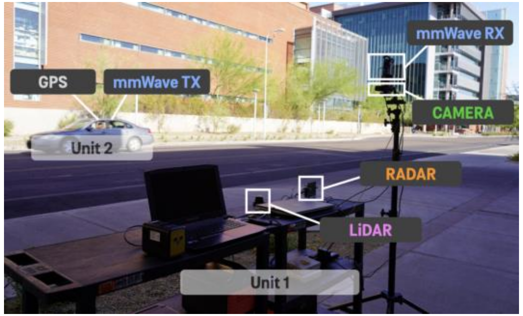
Outline

- Problem Statement
- Multi-Modal Sensing Data Preprocessing
- Deep Learning for Beam Prediction
- Experimental Results and Discussions
- Conclusions and Future Works

Problem Statement

- Communications beyond 5G
  - High frequency (mmWave) and narrow beams
    - Boost capacity, increase SINR, reduce energy
  - Challenges in mobility and beam management
    - Propagation loss, high speed, high reliability
- Sensing-assisted beam prediction
  - Beam prediction from multi-modality sensors
    - 5-instance camera, LiDAR, radar + 2-instance GPS
    - Predict beam with maximum uplink received power
  - GPS: high latency, energy, interruption issues
  - Sensors: environment (blockage, reflection), location



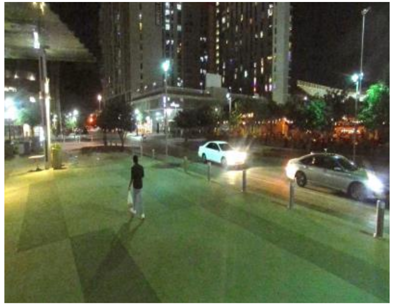


Problem Statement

- Challenges of sensing for beam prediction
  - Data from different time, location, sampling rate
    - Generalization to unseen scenario than training
  - Fusing 3D LiDAR, radar, 2D camera, 1D GPS data
    - Multiple static and mobile objects without labels
    - Misaligned viewing angles of camera, LiDAR, GPS
- Distance base accuracy of top 3 beams
  - Distance to ground-truth beams
    - $Y_K = 1 - \frac{1}{N}\sum_{n=1}^{N} \min_{1\le k\le K} \min\left(\frac{|\hat{y}_{n,k}-y_n|}{5}, 1\right)$
  - Adjacent beams serve better connection
    - $|\hat{y}_{n,k} - y_n|$<=4

Sensing Data Preprocessing

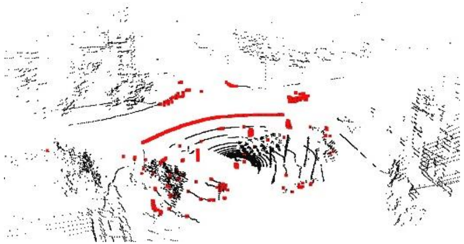
- Camera data:
  - Difficult to recognize targeted user from other mobile agents and backgrounds
  - Enhancing the brightness
    - MIRNet: lighten night scenarios 33, 34
  - Semantic segmentation
    - PDNet: highlight vehicle from background
  - Background masking
    - Blackout background and retain street
  - Guide deep learning model to focus on regions and objects of interests



Sensing Data Preprocessing

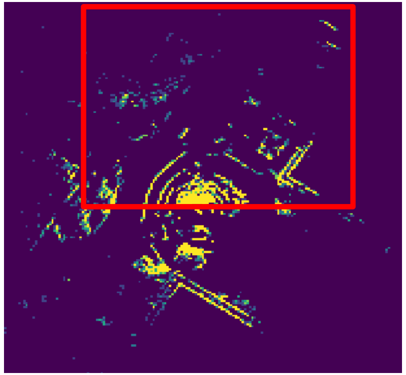
- LiDAR data:
  - Bird Eye View (BEV) projection
    - Discretize ROI into grid cells
    - Encode height, intensity per cell
    - Preserve point-cloud structure in 2D
    - Learn with CNN, less computation
  - Custom Field of View (FoV)
    - Crop BEV to align FoV with camera
  - Filtering backgrounds
    - Filter static points by moving average

3D Point-Cloud



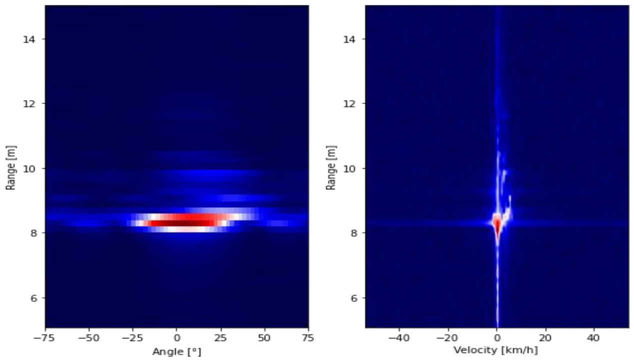
Bird-Eye-View

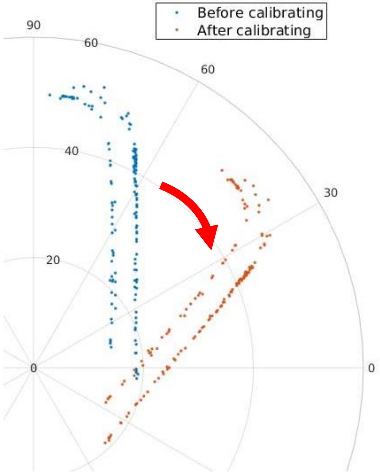
Field-of-View



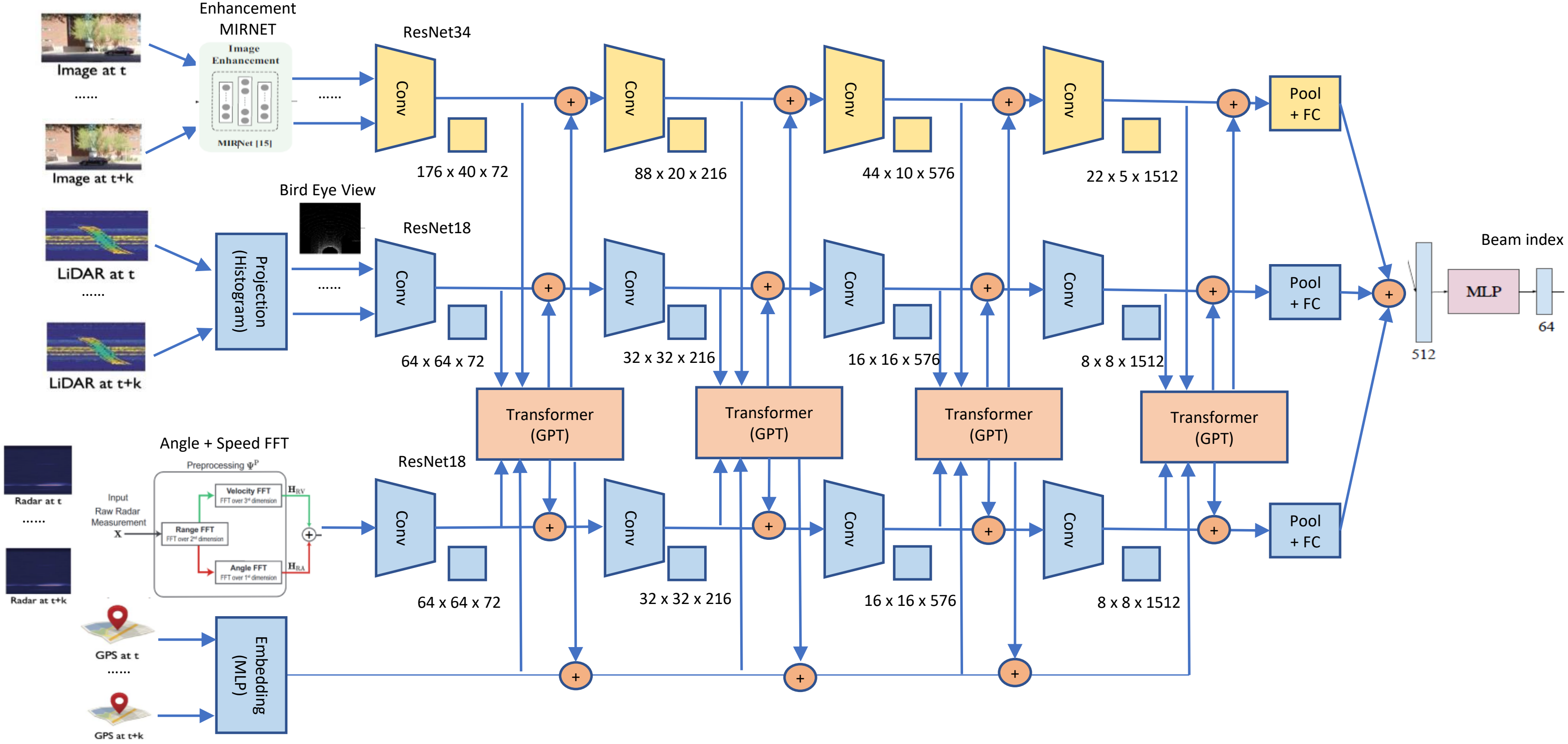
Sensing Data Preprocessing

- Radar data:
  - 2D Fourier transform to produce range-angle and range-velocity maps
  - Reliable speed information without impacts from the environments
- GPS data:
  - Min-max normalization
    - Produce UE relative coordinates $(\Delta x, \Delta y)$ with refer to BS and divide maximum value
  - Calibrated angle normalization
    - Position zero-degree coordinate to the central pixel of images in all scenarios





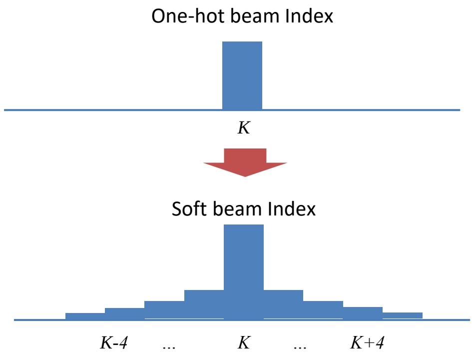
Deep Learning for Beam Prediction

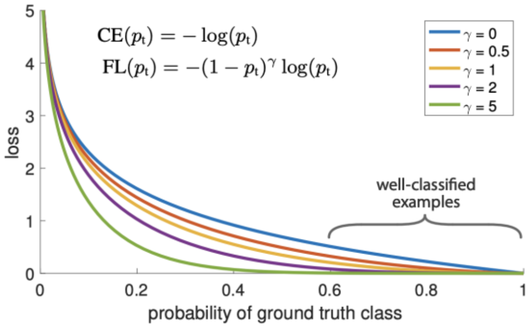


Transformer-based Multi-Modal Sensing assisted Beam Prediction Model

Deep Learning for Beam Prediction

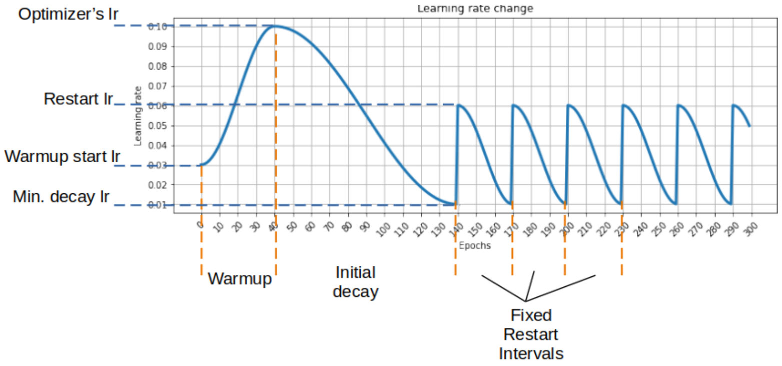
- Soft beam index
  - Change one-hot index to Gaussian distribution
  - Match CE loss function with DBA scores
- Foal loss
  - Modulating factor focus training on hard examples
  - Solve data imbalance between scenarios and class
- Data augmentation
  - Change image brightness, contrast, gamma, hue, saturation, sharpness, blurring
  - Add random noise and downsample LiDAR, radar



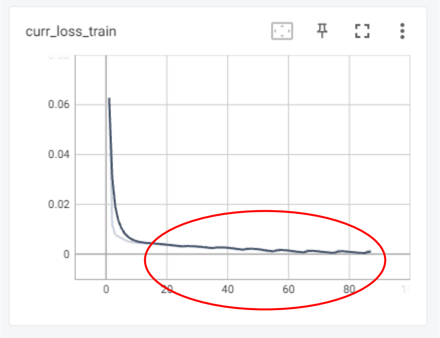


Deep Learning for Beam Prediction

- Cyclic cosine decay schedular
    - Stabilize convergence in training
    - Gradually reduce SGD momentum



- Exponential moving average
    - Improve model robustness
    - Reduce last n step fluctuations



$$\theta_n = \theta_1 - \sum_{n=1}^{n-1} g_i$$

$$\theta_n = \theta_1 - \sum_{n=1}^{n-1} (1 - \alpha^{n-i}) g_i$$

Experimental Results

- Single 5th timestamp
  - High score in trained scenarios
  - Fine tune improves performance
- Multiple timestamps
  - Focal loss reduce imbalance impact
  - GPS angle calibrate improves s31
  - EMA enhance general robustness
  - LiDAR FoV calibrate perform best

TABLE I: DBA score on test dataset of developed schemes

Test	Base	Enhance	Overall	31	32	33	34
A		Timestamp 5 Image enhance Radar angle	0.4618	0.1147	0.6864	0.7848	0.8188
B	A	Fine tune 31	0.5891	0.4718	0.6222	0.6933	0.7328
C	A	Timestamp 1 to 5 Radar velocity Focal loss Cosine decay LR Soft beam index	0.5989	0.4509	0.6852	0.7538	0.7369
D	C	GPS angle norm Data augment	0.5997	0.4713	0.7000	0.7424	0.6997
E	D	EMA	0.6325	0.4760	0.7123	0.7819	0.7985
F	D	LiDAR FoV	**0.6671**	**0.5331**	**0.7173**	**0.7910**	**0.8209**

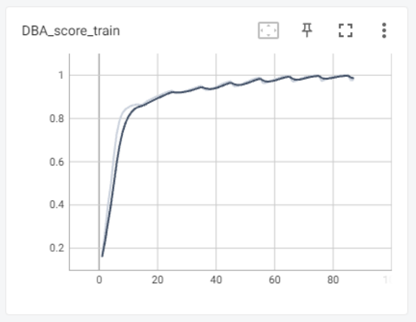
Experimental Results

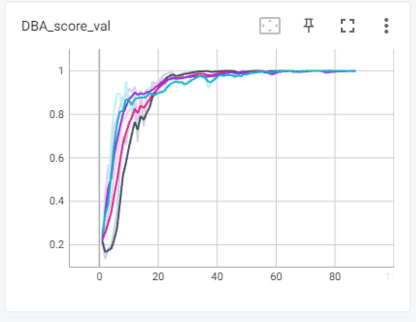
- Background reduction
  - Reduced score when filter, mask and segment on LiDAR and image
  - Visual sensing provide gain from environment information
- Convergence performance
  - Converges at 80 epochs, in all datasets and scenarios
  - EMA reduce fluctuation impacts

TABLE II: DBA score on test dataset of experimental preprocessing

Test	Base	Enhance	Overall	31	32	33	34
G	F	LiDAR filter	0.6398	0.4856	0.7000	0.7914	0.8061
H	G	EMA	0.6458	0.5347	0.6951	0.7505	0.7679
I*	F	Image segment Image mask	0.6298	0.4709	0.7284	0.7810	0.7684
J*	I	EMA	0.6433	0.4947	0.7506	0.7890	0.7837

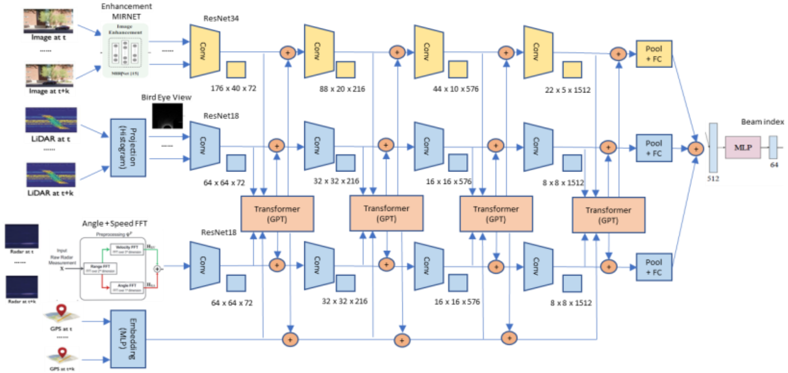
* No image enhancement in scenario 33 and 34.





Conclusion

- Contribution
  - Transformer deep learning for beam prediction
  - Preprocess sequential multi-modal sensor data
  - Generalize to various scenario and applications
- Advantage
  - Tailorable model size, data sequences, modalities
  - Robust in extreme environments: fog, rain, cloud
  - Diverse devices and sensors in wireless network



- Enhancement
  - Contrastive learning improve generalization
  - Semi-supervise learning reduce labeling needs
  - Feature learning improve multi-modal abstraction
- Extension
  - Beam, power, resource management, RIS
  - Sensing, localization, trajectory prediction
  - Collaborative control vehicle, robot, traffic

Technology Innovation Institute

Thank you