Transformer Multi-Modal Deep Learning for Sensing assisted Beam Prediction

ITU AI/ML in 5G Grand Challenge 2022

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

Thanks to Kebin Wu

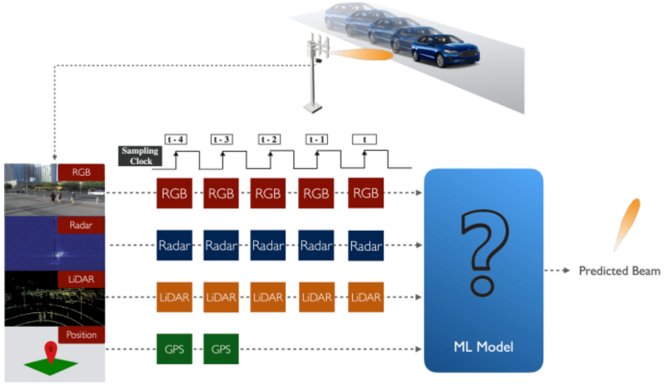
*Team members contributed equally to this work.

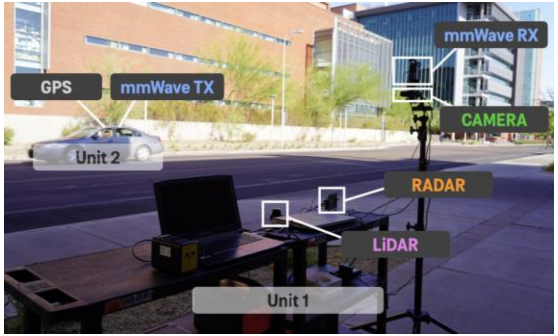
Outline

- 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
  - Exploiting high frequency with narrow beams
    - Boost system capacity and increase SINR
  - Challenges in mobility and beam management
    - Large propagation loss and fast changing channel
    - Require ultra-high reliability and low energy
- Sensing assisted beam prediction
  - Integrated sensing and communications for 6G
    - Sensing in network: camera, LiDAR, radar, GPS
  - Improve beam prediction by multi-modal sensors
    - Information of mobility, blockage, reflection



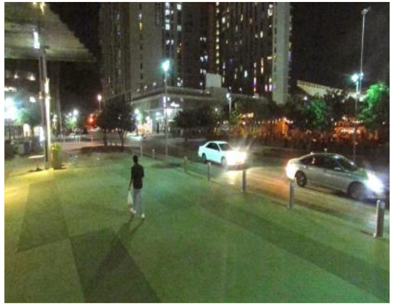


Problem Statement

- Multi-modal sensing data in different scenarios
    - Measured in different time, location, sampling rate
    - Multiple static and mobile objects without labels
    - Generalization to unseen scenario than training
    - Misaligned viewing angles of camera, LiDAR, GPS
    - Fusing 3D LiDAR, radar, 2D camera, 1D GPS data
- Distance base accuracy of top 3 beams
    - DBA-Score: $Y_K = 1 - \frac{1}{N}\sum_{n=1}^{N} \min_{1 \le k \le K} \min\left(\frac{|\hat{y}_{n,k} - y_n|}{\Delta}\right)$
        - Distance to ground-truth top k beam (by receive power)
    - Adjacent beam with high score give better connection

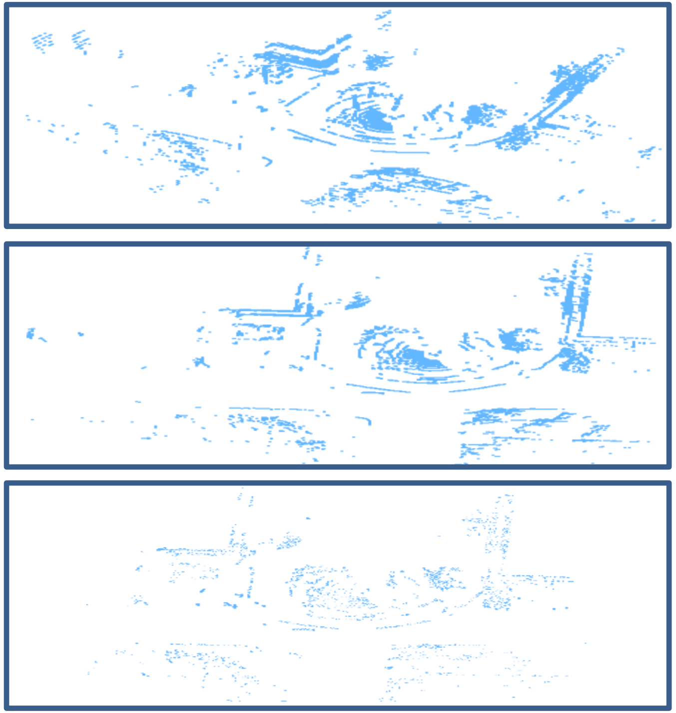
Sensing Data Preprocessing

- 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: lighten vehicle from background
  - Background masking
    - Blackout background and retain street
  - Guide deep learning model to focus on regions and objects of interests



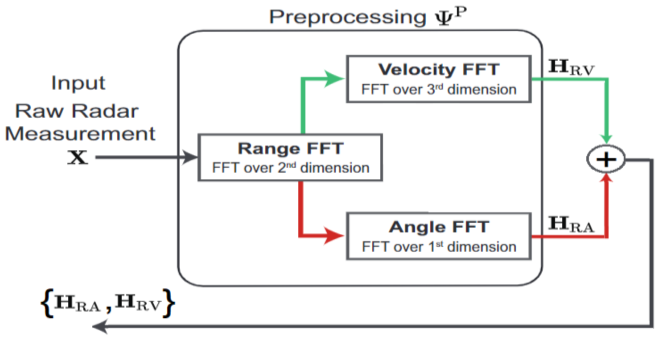
Sensing Data Preprocessing

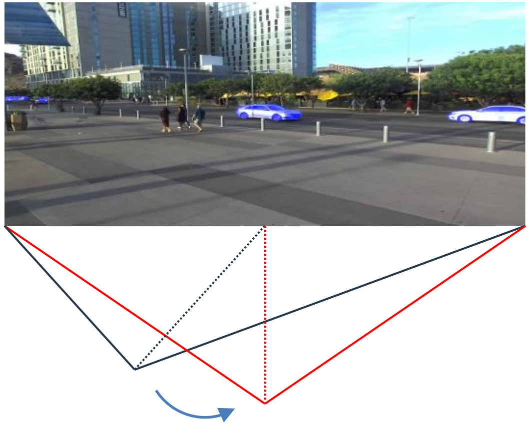
- LiDAR data:
  - Filtering backgrounds
    - Remove static points by moving average calculation cross the dataset
  - Bird Eye View projection
    - Discretize region of interest into grid cells
    - Encode height and intensity per maximum value in each cell, project to 2D channels
    - Produce image-like representation that preserve structure with less computation
  - Custom Field of View
    - Crop BEV to align FoV with camera



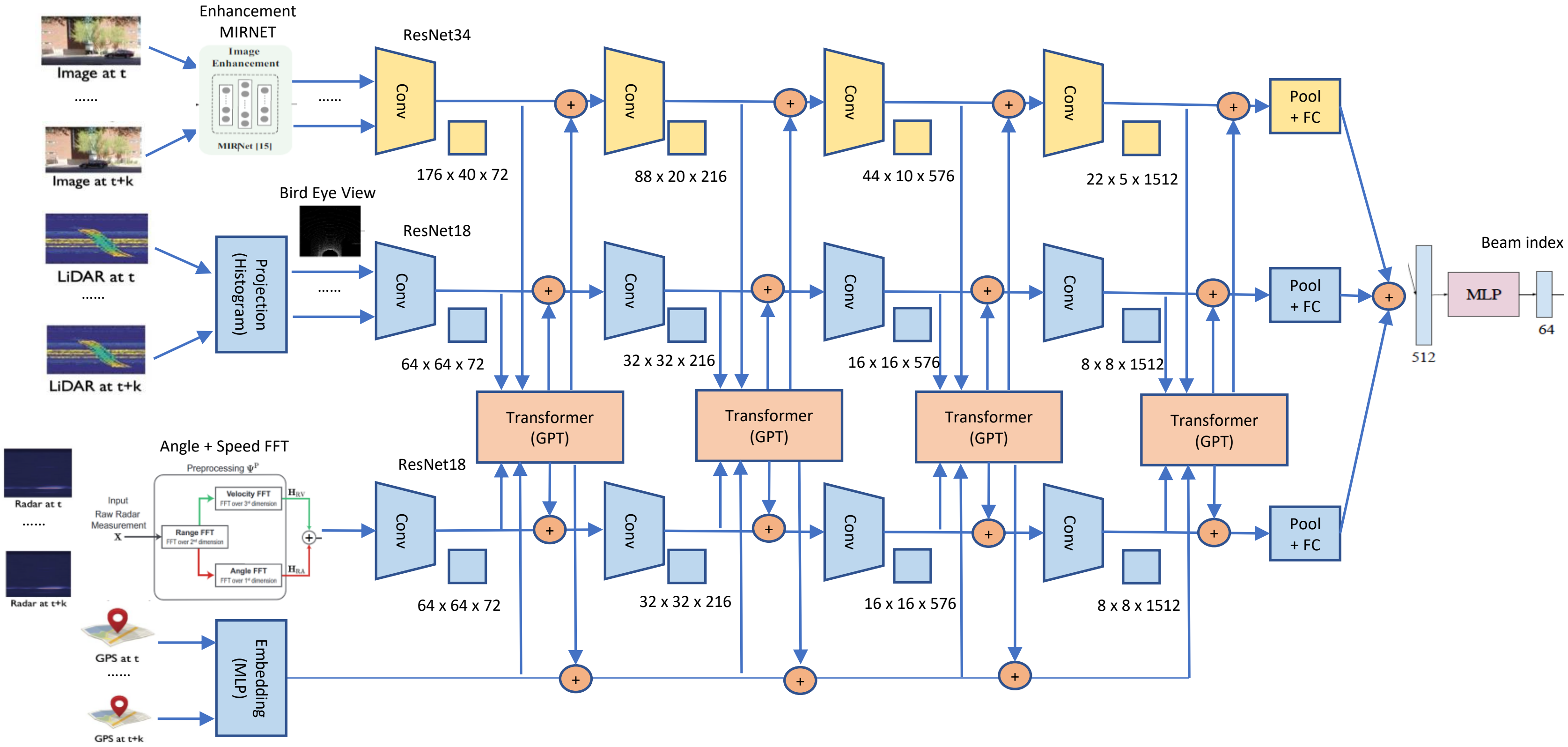
Sensing Data Preprocessing

- Radar data:
  - Perform 2D Fourier transform to produce range-angle and range-velocity maps
  - Produce 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



Deep Learning for Beam Prediction

- Soft beam index
  - Change one-hot best beam index to Gaussian distribution with cutoff at 5
  - Soften the cross-entropy loss function to match the DBA score function
- Foal loss
  - Apply modulating factor to reduce loss contribution from easy examples
  - Solve data imbalance between scenarios by focus training on hard examples
- Data augmentation
  - Change image brightness, contrast, gamma, hue, saturation, sharpness, blurring
  - Add random noise to LiDAR, radar and random down sampling LiDAR data

Deep Learning for Beam Prediction

- Cosine decay schedular
  - Start training with high learning rate, decrease to minimum then restart
  - Stabilize convergence after long run and preserve stochastic optimization
- Exponential moving average
  - Maintain parameter values during training and store EMA of last n steps
  - Improve model robustness by avoiding fluctuation after long training

Experimental Results

- Single 5th timestamp
  - High score in trained scenarios
  - Fine tune improves generalization
- Multiple timestamps
  - Improve score cross scenarios with imbalanced data using focal loss
  - GPS angle calibration contributes to improvement in unseen scenario
  - EMA largely improves score in all scenarios by making model robust
  - LiDAR FoV calibration contribute to best score by alignment of scenes

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**

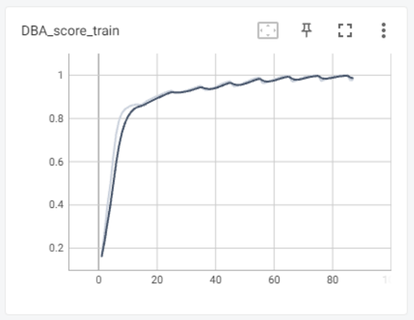
Experimental Results

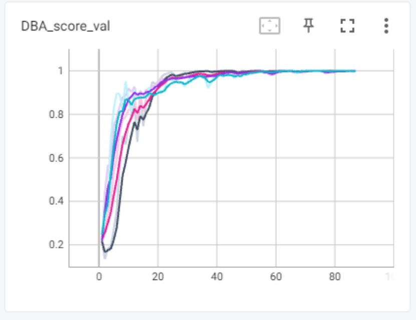
- Background reduction
    - Reduced score when filter, mask and segment on LiDAR and image
    - Visionary information benefit beam predicting by learning propagation environment and relative position
    - EMA enhance model robustness
- Convergence performance
    - DBA score converges at 80 epochs, both training and validation datasets in all the 4 scenarios

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.





Conclusions and future works

- Innovative contribution
  - Transformer multi-modal learning framework for sensing aid communication
  - Novel sensor data preprocessing and training method for communications
  - Improve beam prediction accuracy to 66.71%, and 55.31% in new scenario
- Future works
  - Generalizing framework to wider communication, sensing, control applications
  - Enhance generalization with semi-supervise learning and contrastive learning

Technology Innovation Institute

Thank you