## Basic Information

Chinese name: 成泽森

Pinyin: Cheng Ze Sen

English name: akko

Place of Birth: Hunan Yongzhou (湖南永州)

Birth Date: 1999.11.10

Major: Electronic Science and Technology

CET6: 522 (Listen:182 Read:211 Writing:129)

Scores: 91.12 GPA: 3.92 Rank: 1/88

I’m an undergraduate student of South China University of Technology (华南理工大学)

Now, I’m fatter…

## Awards

1. National Scholarship (2017 ~ 2018) [Intellectual education: 1st, Comprehensive assessment: 1st]
2. The Second Prize Scholarship (2018 ~ 2019) [Intellectual education: 1st, Comprehensive assessment: 3st]
3. Cstro 16th Conference -- AI and Big Data Part -- ROI Segmentation Challenge -- Head and Neck OAR track -- Champion
4. National College Students Mathematical Modeling Competition Provincial Third Prize
5. Taidi Cup 8th Data Mining Challenge Second Prize

## Research Experiences

1. Multi-media Emotion Analysis (SRP project From 2019.6 to 2020.4)
2. Sun Yat-sen University Cancer Center Research Assistant (From 2019.8 to now)
3. SCUT-Robot lab Member of Vision Group (From 2019.9 to now)

Detailed:

1. In this project, I’m mainly in charge of implementation. We try to use neural network to extract the emotion feature representation from Multi-media data like image, video, audio etc. Then we use regression and classification model to model the emotion state.

2. I’m the research assistant of MIACA research group Sun Yat-sen University Cancer Center. I focus on medical image segmentation especially multi-modality medical image segmentation. Besides, I am also interested in medical image registration and medical image generation.

3. I’m a member of Vision Group. I’m mainly in charge of base station. Base station can use object detection, object track and motion forecasting to strategy calculation so that it can provide some strategy support for other robots.

## Publications

《基于深度学习的医学图像配准综述》

p.s. [**rejected**]

《Auto Segmentation of Pelvic OARs On MRI Multi-Sequence Using A Fused-Unet》

p.s. [**received**] (AAPM 2020 **oral presentation** & **BLUE RIBBON ePOSTER**)

《A Novel Hybrid Network for H&N Organs At Risk Segmentation》

p.s. [**received**] (ICBIP 2020 **oral presentation**) -- **Corresponding SCI Under Review**

《Attention V-Net: A Residual U-Net with Attention Gate Block for Lung Organ At Risk Segmentation》

p.s. [**received**] (CSAE 2020) -- **Corresponding SCI Under Review**

《ZigZag U-Net: Multi-stage medical segmentation network》

p.s. [**to be submitted**] (inspired by CB-Net)

《Defect of insulator detection based on region》

p.s. [**to be submitted**] (The entry paper of taidi cup data mining challenge)

## Research Interest

Because of my past research, I may be interested in medical image segmentation, object detection and object track. Actually, I’m not limited to specific computer vision tasks. I’m interested in computer vision tasks which are meaningful and practical. For instance, the task about the tumors and organs at risk automatic contouring on MRI Multi-sequence. MRI Multi-sequence tumors and organs at risk manual contouring is really time consuming and clinical knowledge intensive. So, we use Multi-modality model to perform MRI Multi-sequence automatic contouring so that the model largely reduces the contouring time. By doing like so, we can largely ease the pressure of clinical doctor and the time of whole radiotherapy pipeline can be simplified to a great extent.

## The implementation of VectorNet

1. Representing trajectories and maps
2. Constructing the polyline subgraphs
3. Global graph for high-order interactions

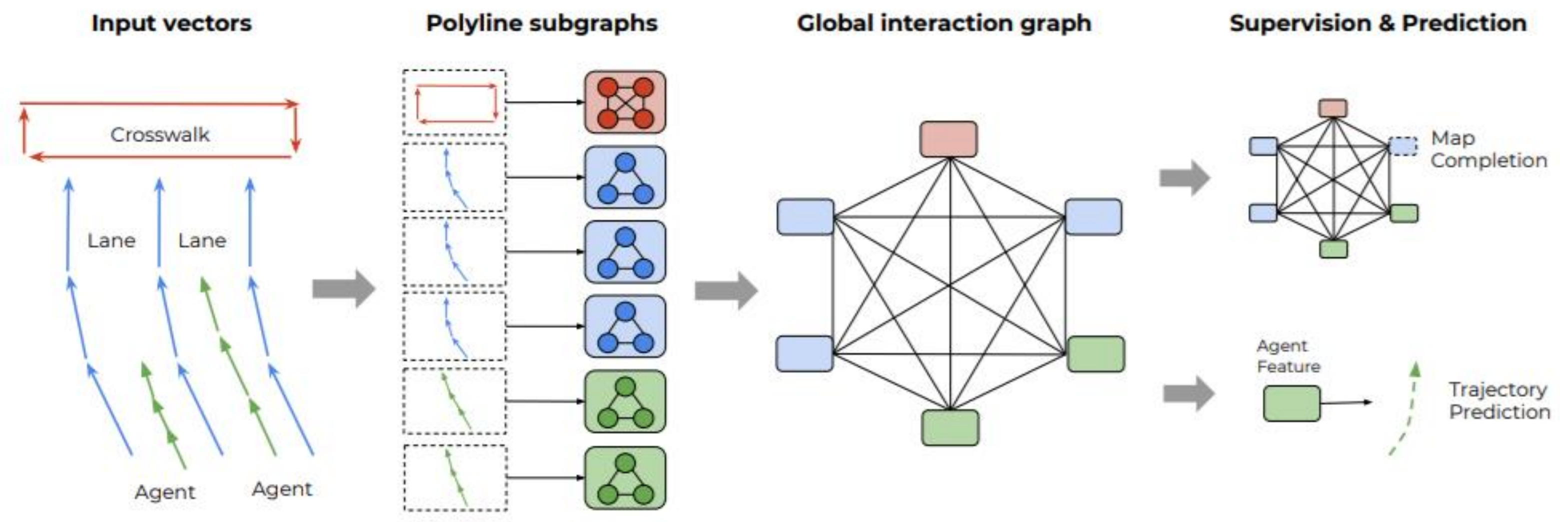


**Figure 1. The total workflow of VectorNet**

## Representing trajectories and maps

There are only the agent trajectory and lane needed to represent according to the requirements. In order to represent lane better, I use average lane width to calculate lane two-side edge lines. Besides, I split the trajectory into observed trajectory [0s, 2s) and future trajectory [2s, 5s).

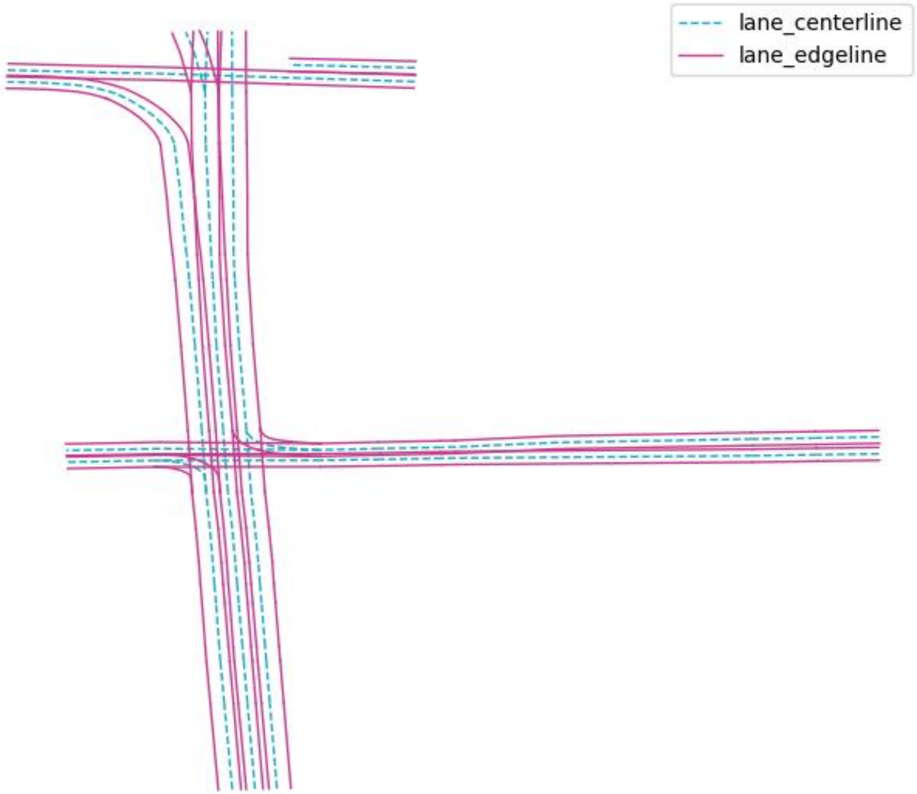


**Figure 2. lane centerline and edgeline representations**

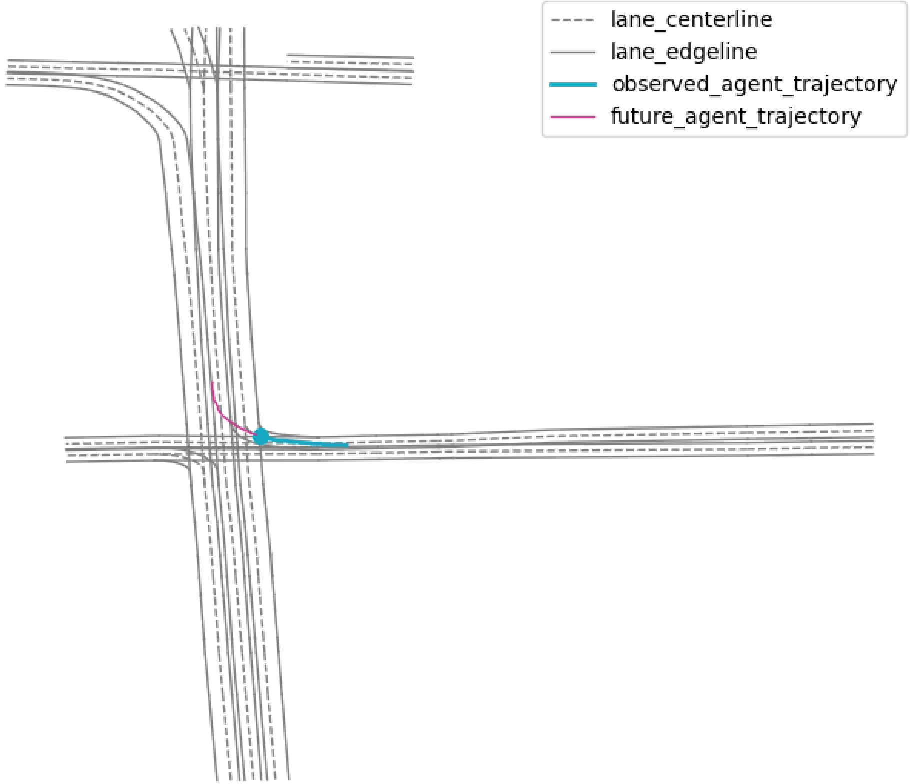


**Figure 3. lane centerline and edgeline representations, agent observed trajectory, agent future agent trajectory**

## Representing trajectories and maps

Because all of the lanes and trajectories are represented by point sets initially. So, we need to incorporate neighboring two points to vectorize these points and construct vector sets. Then we keep vector sets and remove point sets. Because the future trajectory sample points are our predict target, so we keep them.

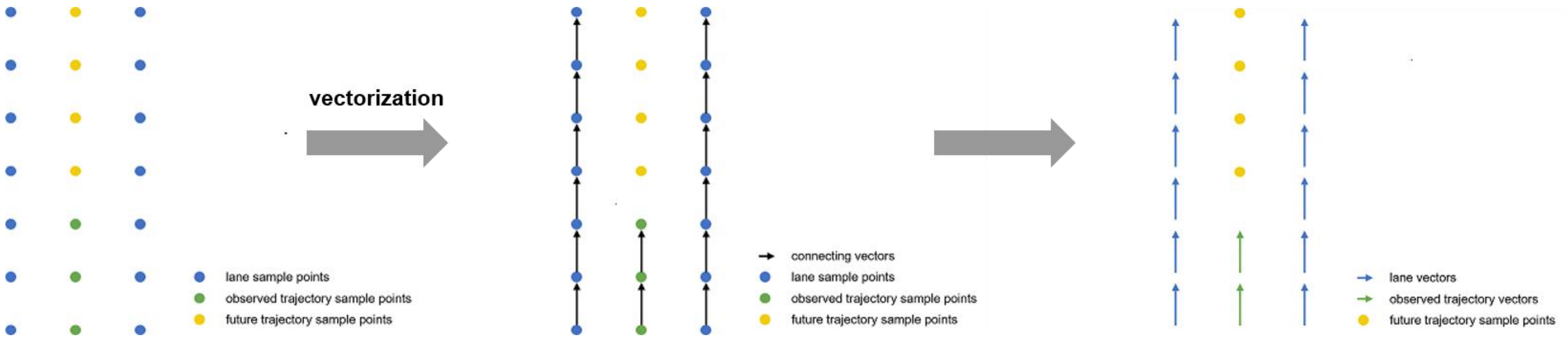


**Figure 4. The process of vectorization of agent and observed trajectory sample points**

## Constructing the polyline subgraphs

This stage is actually the embedding stage. According to the paper, one vector set represents a polyline. As stated before, one lane can be seen as a polyline and the observed trajectory can also be seen as a polyline. So, we will convert these vector sets to polyline level features by constructing the polyline subgraphs.

$\{v_1^0, v_2^0, \cdots\cdots, v_i^0, \cdots\cdots, v_{p-1}^0, v_p^0\}$ representing raw vector sets.
$\{v_1^0, v_2^0, \cdots\cdots, v_i^0, \cdots\cdots, v_{p-1}^0, v_p^0\}$ representing vector sets which has been operated $l$ time.

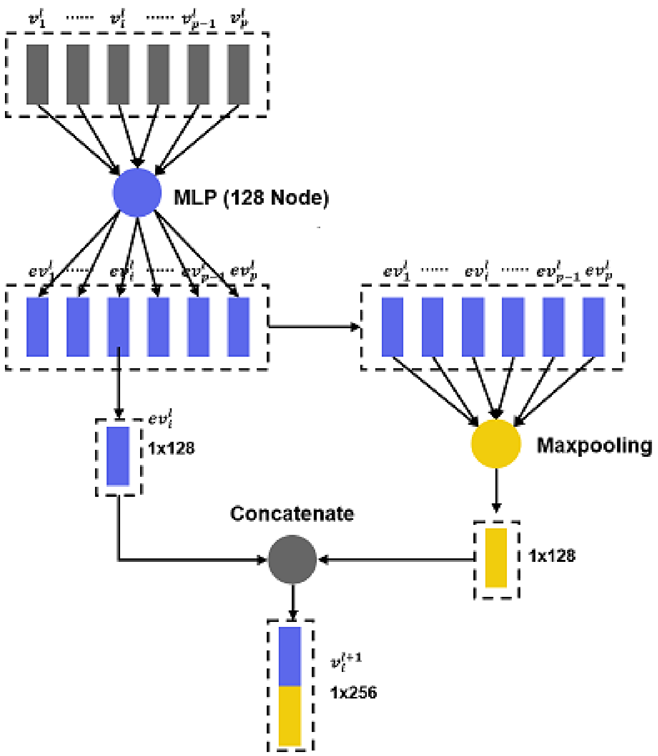


**Figure 5. The detailed operation flow about how to get $v_i^{l+1}$ from $\{v_1^l, v_2^l, \cdots\cdots, v_i^l, \cdots\cdots, v_{p-1}^l, v_p^l\}$**

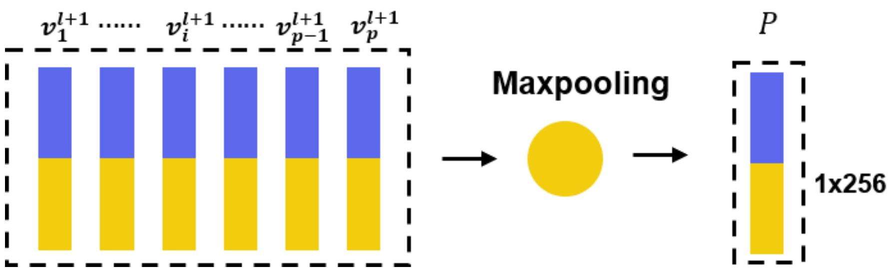


**Figure 6. The detail operation flow of aggregating $\{v_1^{l+1}, v_2^{l+1}, \cdots\cdots, v_i^{l+1}, \cdots\cdots, v_{p-1}^{l+1}, v_p^{l+1}\}$ to polyline level feature $P$**

## Global graph for high-order interactions

We refer to the self-attention mechanism and analyze the detailed calculation flow of self-attention mechanism. The global graph neural network is related to the multi-head attention to a great extent. They copy raw feature map into three copies and call them query, key, value. Their aim is to use matrix multiplication to get weight matrix from query and key. Then using matrix multiplication to assign weight matrix on value matrix.

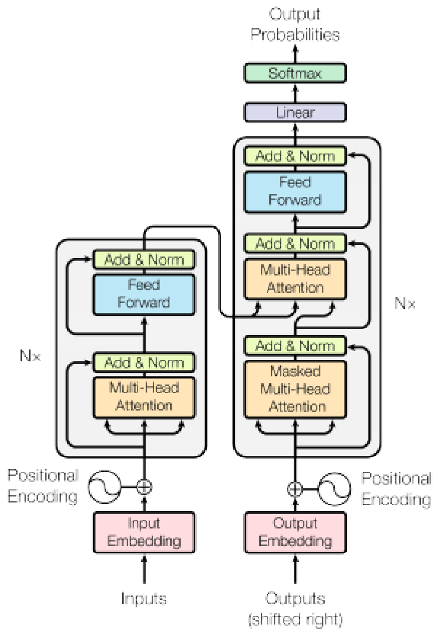


Figure 7. The total pipeline of self-attention mechanism

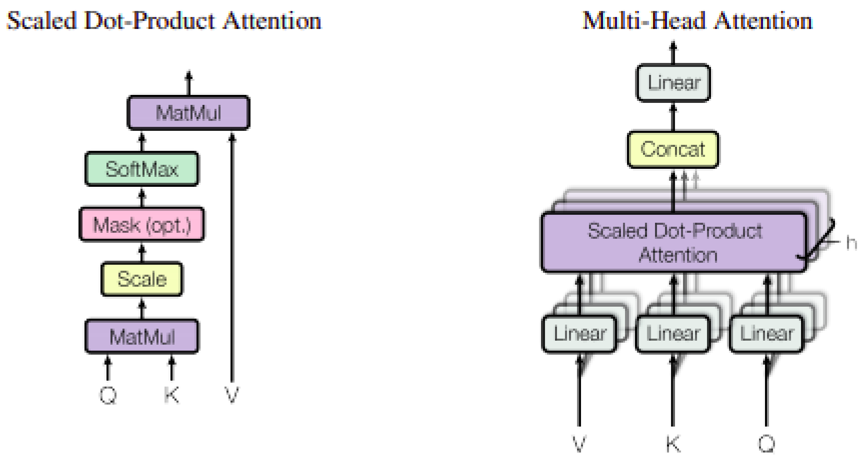


Figure 8. The detailed architecture of multi-head attention which is the reference of global graph of VectorNet

## Global graph for high-order interactions

It seems that it has nothing to do with the graph network. In fact, the matrix multiplication operation is just a type of fully connected interaction between different graph node. The graph node features shape as $[b,\ L^f,\ n^p]$ where $b$ denotes batch size, $L^f$ denotes feature length or the width of global graph, $n^p$ denotes polylines count. After interacting with other graph node, agent polyline feature is decoded by MLP and output agent prediction trajectory points.

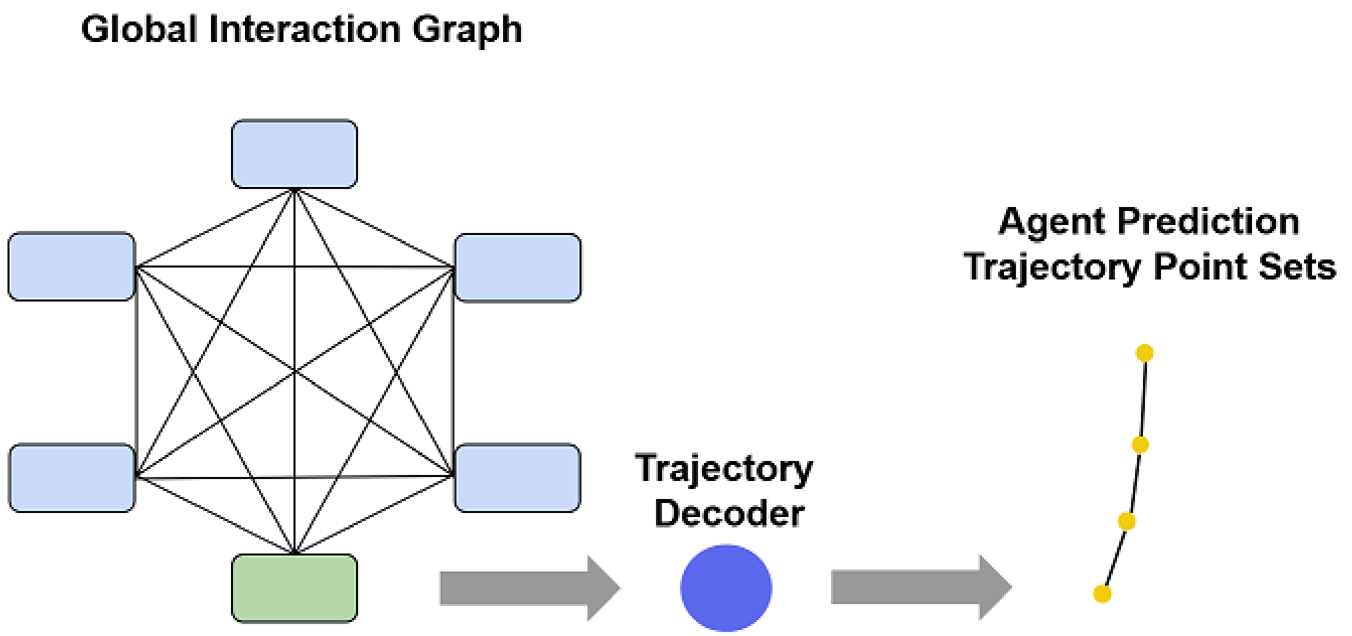


**Figure 9. Global graph interactions and trajectory decoder**

## Other methods

1. Fast and Furious (a end-to-end compound of regular computer vision task)
2. IntentNet (an improved FaF by adding rendered HD map)
3. Rules of the Road (modeling the semantic label by convolution)

Detailed:

1. Firstly, converting the Lidar 3D point clouds as 3D tensor which is a type of regular data which can be encoded by ConvNet. Then stacking 5 timestamps voxelized point clouds as a block. Inputting this block into the ConvNet and the ConvNet will output current and future bounding box.
2. The method of this paper is really similar to the IntentNet. Except inputting voxelized point clouds, they also input rasterized map. Besides, they also perform a new task intention classification and the output format is different with FaF. They directly output the trajectory points.
3. They deliver the scene representation into the network by just inputting rasterized map. Besides, they input the motion state scales of target entity and other entities like velocity, acceleration etc. The network of them will output probabilistic grid map which denote the entity state probability of each position. In fact, they also have other different output format like trajectory points.

## My feeling

Motion forecasting is a very important task in self-driving domain. Because, the autonomous vehicle have to share roads with human drivers who are really stochastic. The main aim of motion forecasting is to avoid the autonomous vehicle crashing with other objects. But the main problem of motion forecasting is that there is not an unified representation which can model the road context and other vehicle motions. Before VectorNet, researchers try to rasterize this information because they really want to use ConvNet to encode them. But VectorNet, I think, is really innovative and efficient. VectorNet is more straightforward to utilize the structed HD map. However, VectorNet also has problem. It may be too idealistic. Because the structed HD map may be difficult to get in real-time. Besides, the observed trajectory of agent is also difficult to acquire in real-time. So there is still a gap to let VectorNet equipped in the autonomous vehicle.

Thanks for listening