AML for multi actor forecasting in 5 minutes

(GenZ Multidimesnional Multitasking)

December 2022

Universiteit Leiden
The Netherlands

Tobias Florin Oberkofler,
Supervisors: Mitra Baratchi, Javier Alonso-Mora, Holger Hoos

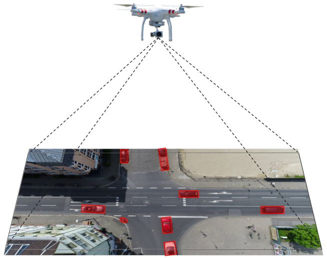
Problem statement

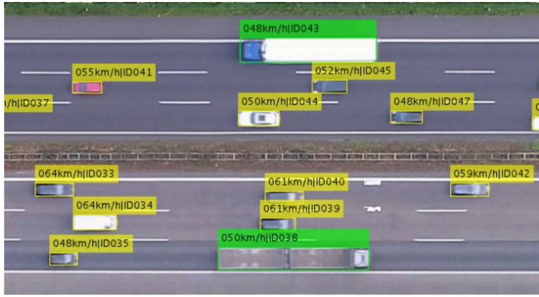
- Multistep Trajectory prediction
  - Temporal & Spatial & (Social) dimension

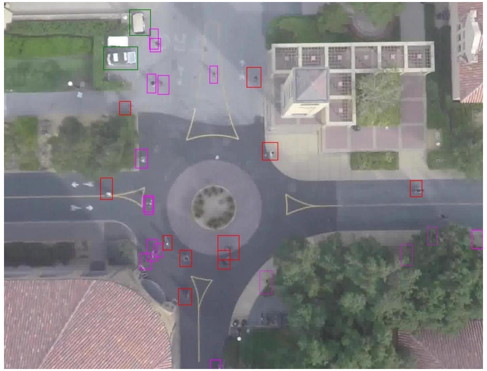
    In the concrete case we consider the (planar) coordinates as attributes along derivatives such as *velocity, acceleration, angle, etc.*
  - *Context and exogenous* variables are also of importance (*epistemic uncertainty and aleatoric variability)*

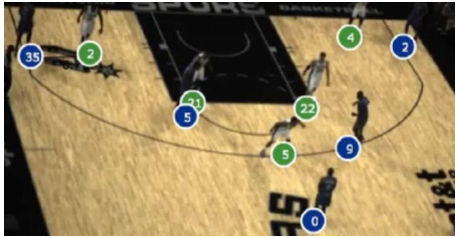
Benchmarks

- **HighD, InD**
  - Vehicle trajectory dataset benchmark
- **OpenTraj, TrajNet++, NBA**
  - Human Trajectory Prediction Dataset Benchmark
  - sport motion-capture









Details and non-standards:

- **2.5Hz**, 5Hz
- **3s obs - 5s pred**, 1.5s obs - 2.5s pred
- **frame** vs second wise prediction

Problem statement

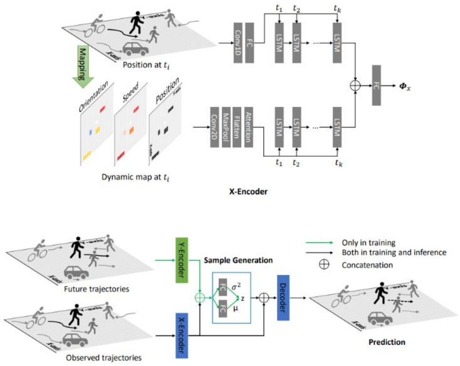
- **Short-term** trajectory prediction:
  - Often it is impossible to know the exact intent of an agent
    - multi-modal prediction instead of the deterministic future $p_\theta(Y_n|X)$
  - Measure average and final displacement error:

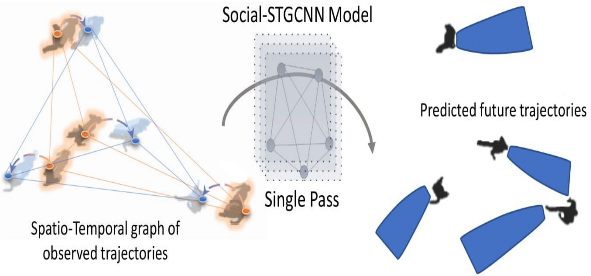
$$\mathrm{ADE}_K = \frac{1}{T} \min_{k=1}^{K} \sum_{t=T_{\text{obs}}+1}^{T_{\text{pred}}} \left\| \hat{\mathbf{y}}_n^{t,(k)} - \mathbf{y}_n^t \right\|^2 \quad \mathrm{FDE}_K = \min_{k=1}^{K} \left\| \hat{\mathbf{y}}_n^{T,(k)} - \mathbf{y}_n^T \right\|^2$$

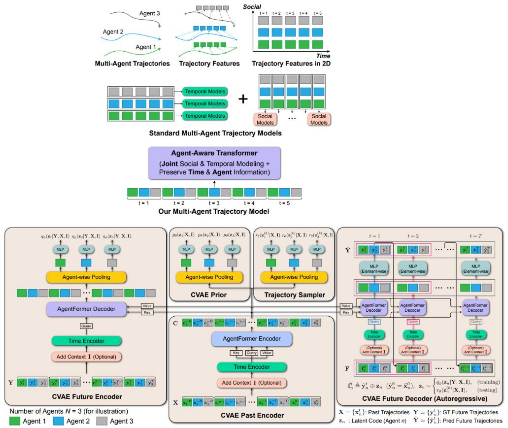
Architectures

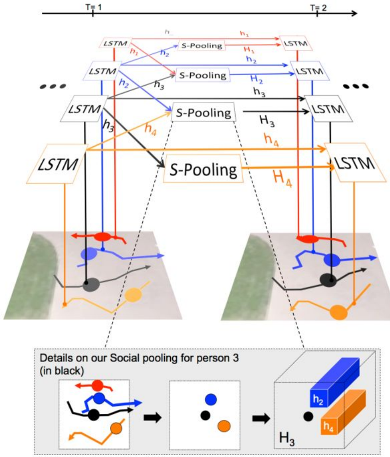
- Rule based baselines:
  - Kalman Filters (cv) [1960?]
  - Social Forces [1995]

- SOTA Neural networks:
  - Cheng et al. 2020 (AMENET)
  - Xu et al. 2022 (SVAE)



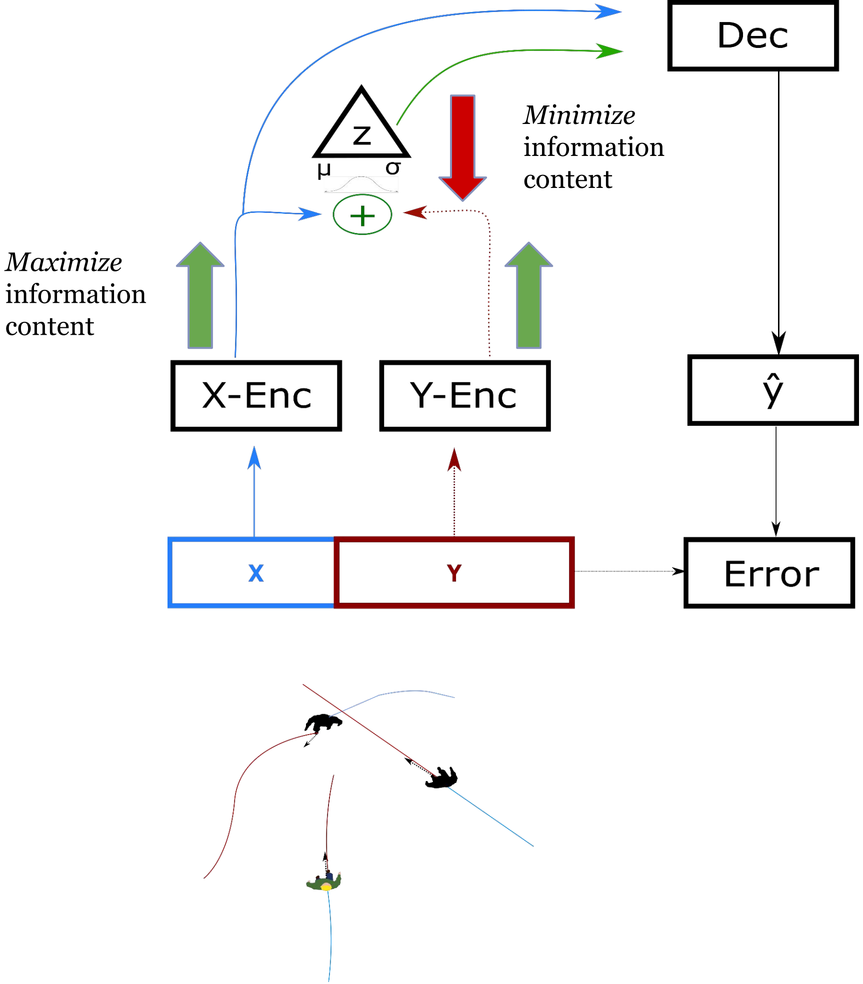






Architecture

- Macro-architecture:
  - Variational Autoencoder (*Recurrent Variational Neural Network*)
- *Encoder for past and future* to extract useful motion and interaction patterns
- *Latent variable* to model epistemic uncertainty
- MSE between y and ŷ
- KL-Divergence: between prior and approximate posterior



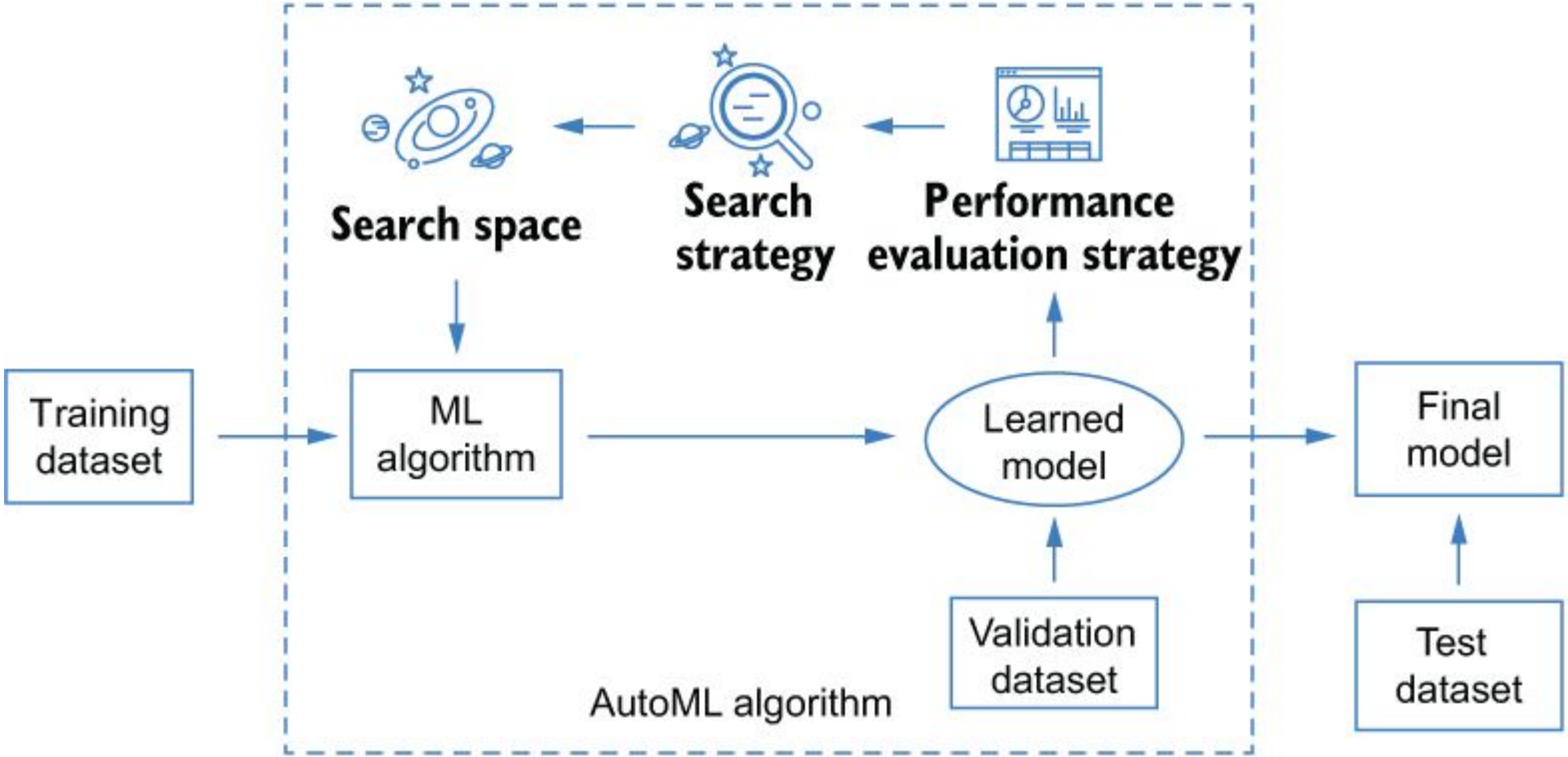
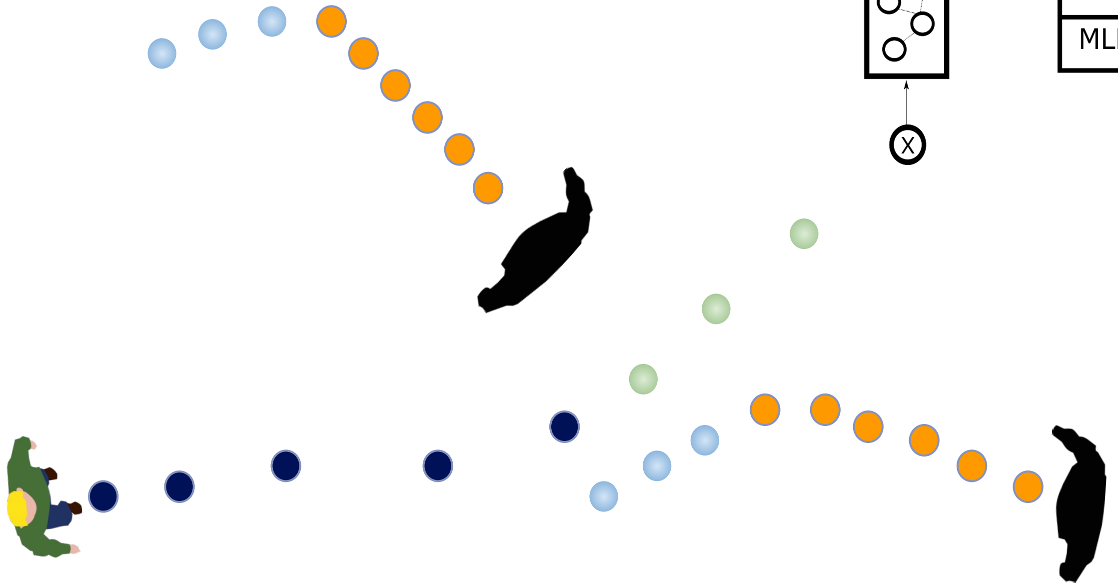
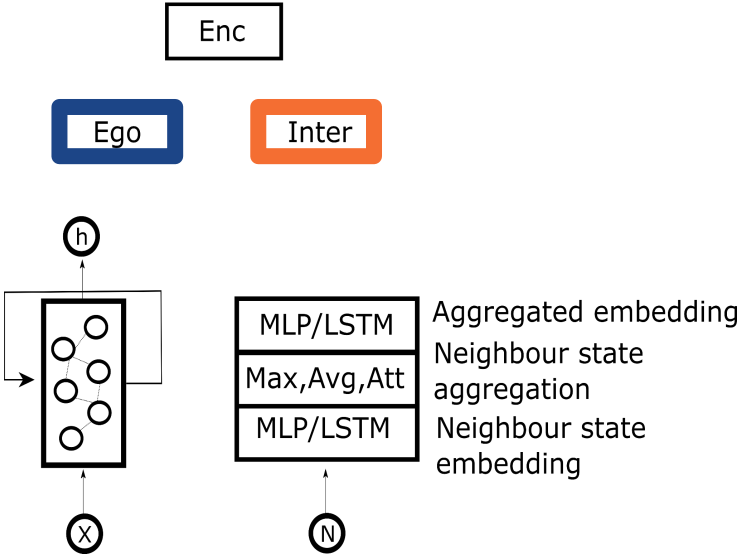


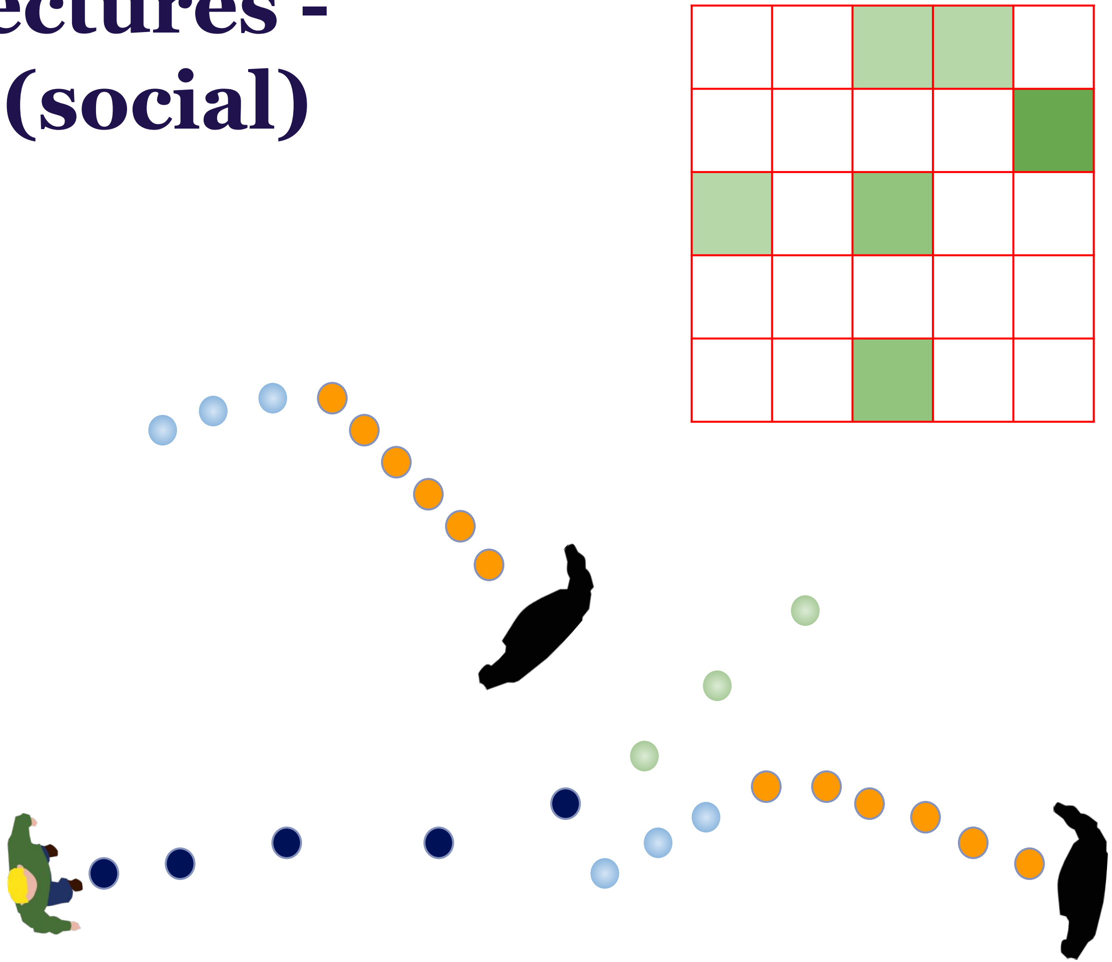
Image source: Automated machine learning in action - song et al.

Architectures - Cell based

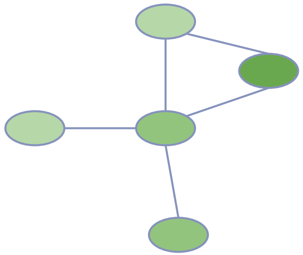


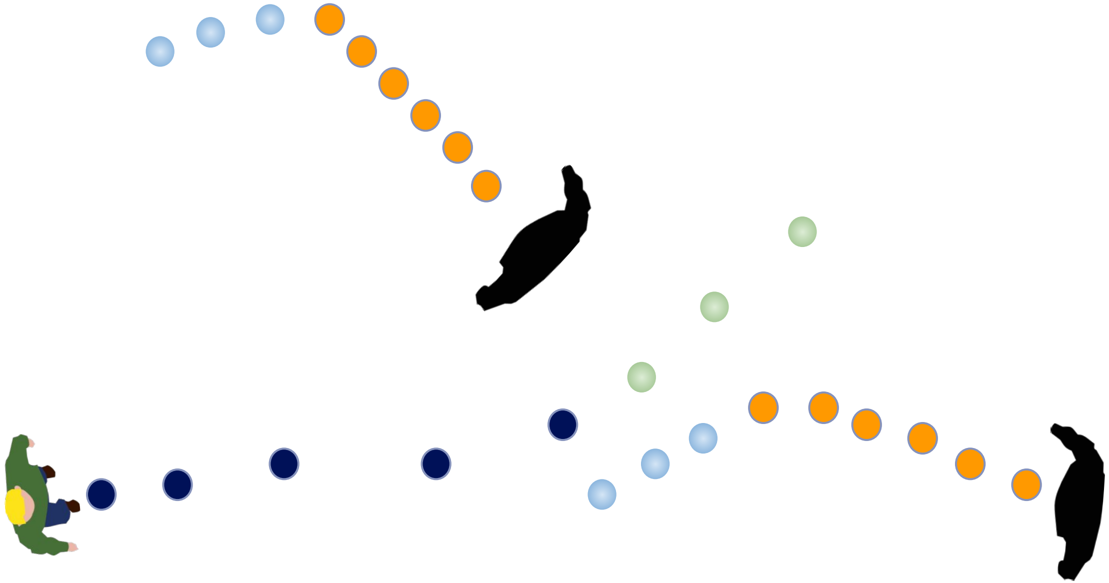


Architectures - Macro (social)



Architectures
- Macro v2





Implementation

- Use (Beta)DARTS for efficient search with gd.
  - Weight sharing within supernet
  - Iteratively update architecture parameters on validation and network parameters on train set
  - regularize params for robustness

- Use NNI as implementation framework: modular and extendable. Large community. Allows intuitive impl.

- Works on GPU and CPU (faster with multiple GPUs)

Experiments

- Establish Baseline performance

- Setup:
  - Trajectory data with *frame,id,x,y*
    extract: vel, acc, dist, rel_vel, rel_acc, rel_angle
  - 8 (3s) obs to 12 (5s) pred split (NBA: 8(1.4s) to 12 (2.5s) split)
  - HP settings for baselines as outlined in papers when trained on most similar dataset (for most the same) or searched in 12 hours TPE optimization investigating lr, bs, cell size, latent dim
  - 5-fold cross-validation
  - Train/Val at 80/20 split
  - Metrics: average displacement, final displacement

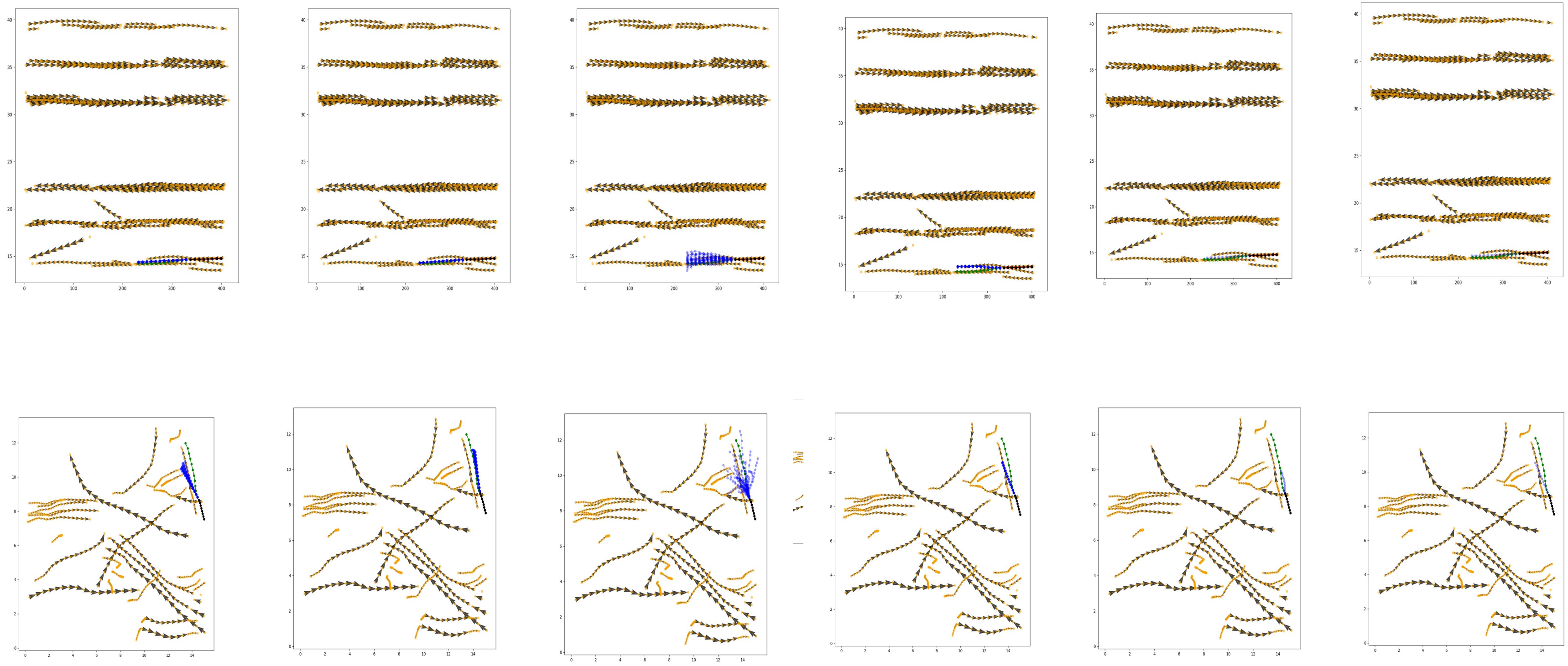
Experiments: Results

Model	# outperformed	Average ranking
	ADE	ADE
SVAE	**16/4**	**2.4**
AMENET	14/5	3.4
AML-MRNN	13/6	3.4
AML-SLSTM	10/8	3.6
AML-Cell	9/7	4.2
CV	8/6	4.4
SF	0/4	6.6

Model	# outperformed	Average ranking
	FDE	FDE
AML-MRNN	**18/6**	**2.2**
SVAE	16/4	2.4
AMENET	14/5	2.5
AML-SLSTM	8/7	4.8
AML-Cell	8/7	4.8
CV	6/7	5.0
SF	2/4	5.2

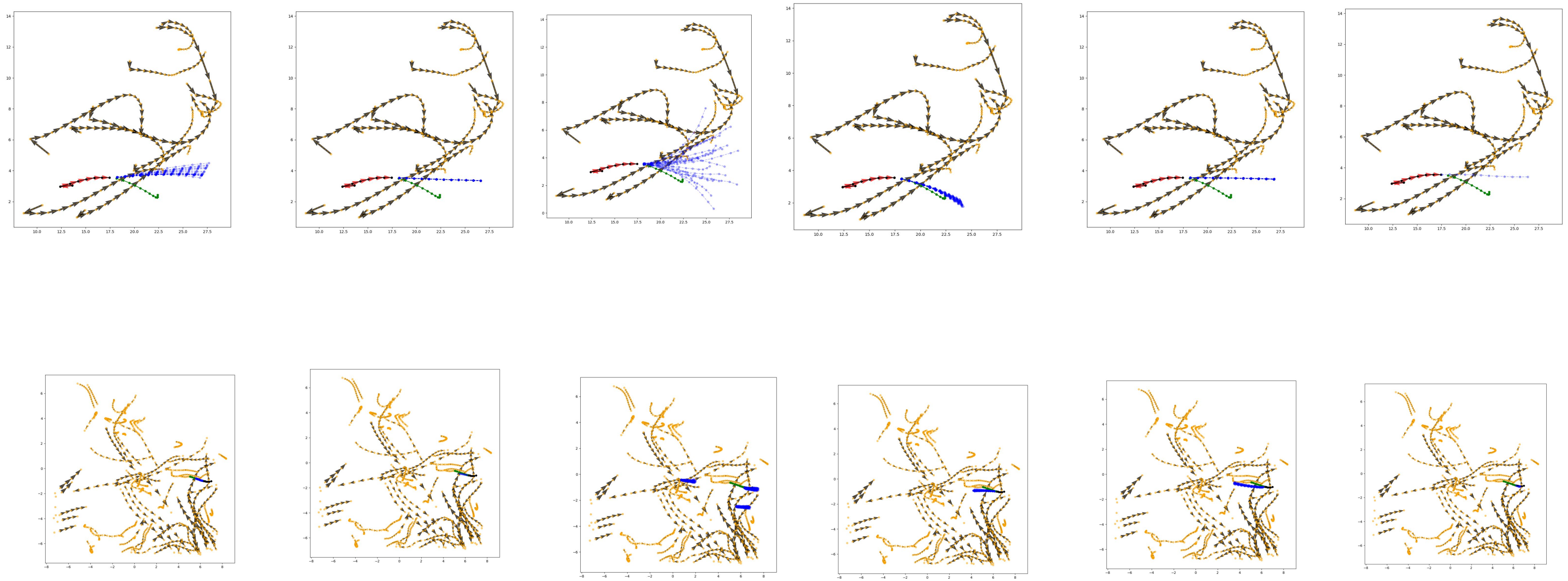
Experiments: Results

AML-MRNN AMENET SVAE AML-SLSTM AML-Cell CV

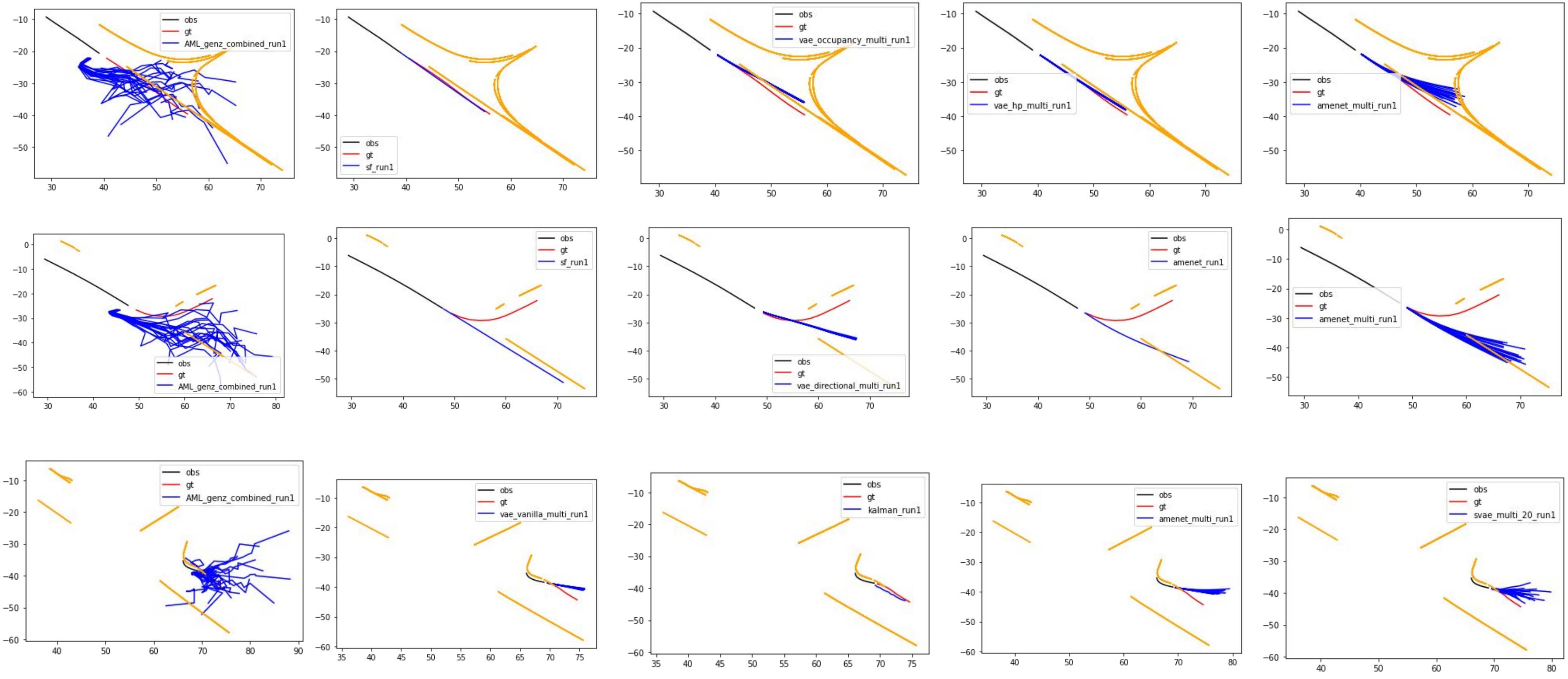


Experiments: Results

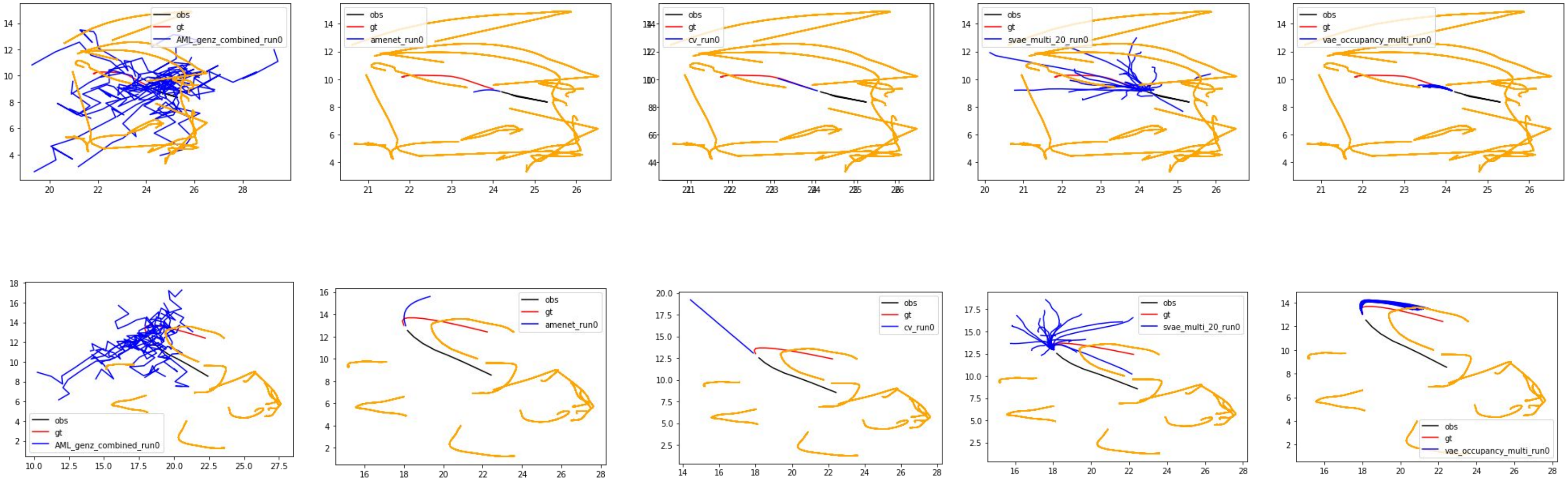
AML-MRNN AMENET SVAE AML-SLSTM AML-Cell CV



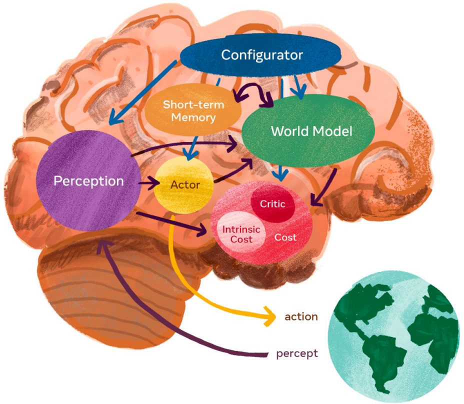
Experiments: Results

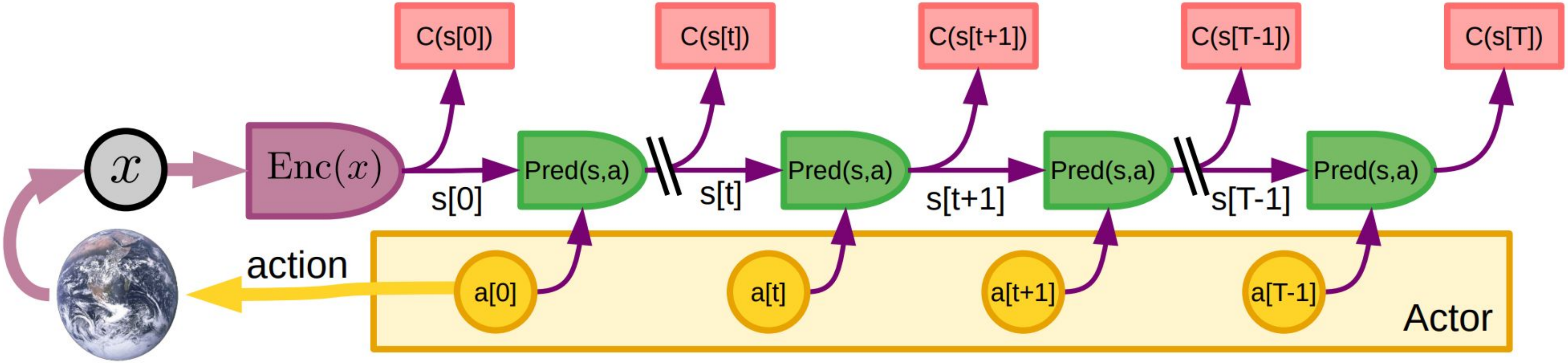


Experiments: Results



Future


Configurator
Short-term Memory
World Model
Perception
Actor
Critic
Intrinsic Cost
Cost
action
percept


C(s[0])
C(s[t])
C(s[t+1])
C(s[T-1])
C(s[T])
x
Enc(x)
Pred(s,a)
Pred(s,a)
Pred(s,a)
Pred(s,a)
s[0]
s[t]
s[t+1]
s[T-1]
action
a[0]
a[t]
a[t+1]
a[T-1]
Actor

SUMMARY

- Propose a new varied benchmark for short-term trajectory prediction
- Large scale baseline results for the benchmark outlining performance under different data properties
- Investigate novel search space for first time NAS with trajectory data and in the context of stochastic macro-structure
- Exploration of real world behaviour
- Discussion of metrics

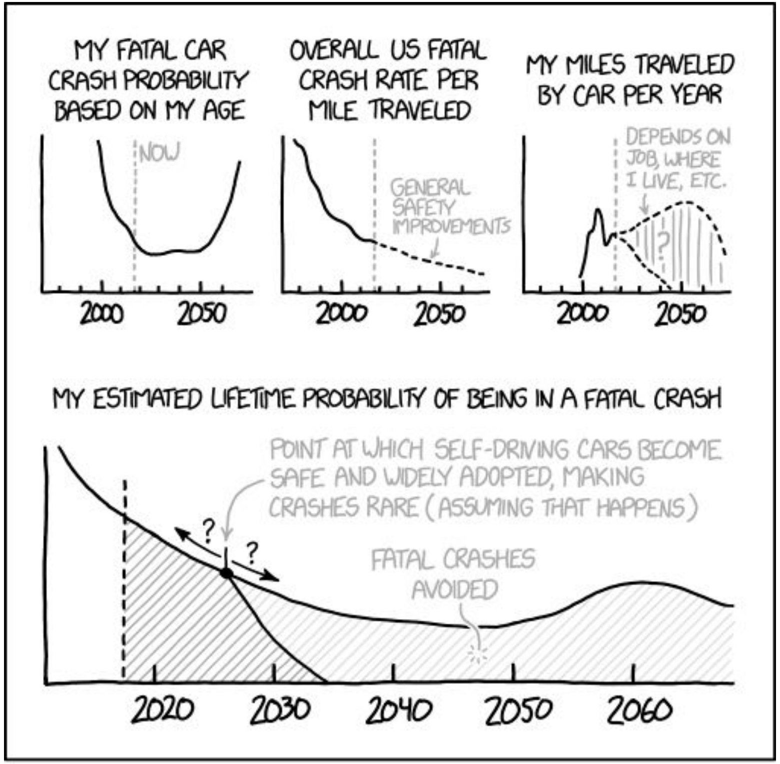
https://github.com/to314as/sospatem

Questions?

Inputs?

Discussion?

https://github.com/to314as



IT FEELS WEIRD TO LOOK AT CAR CRASH STATISTICS AND WONDER WHETHER WE'LL ALL BE ABLE TO STOP DRIVING BEFORE I'M INVOLVED IN A FATAL CRASH.

$$Y = (Y^{t+1}, Y^{t+2}, ..., Y^{t+k})$$

(Beta) DARTS

DARTS: continuous relaxation of the architecture representation to allow efficient search with gd.

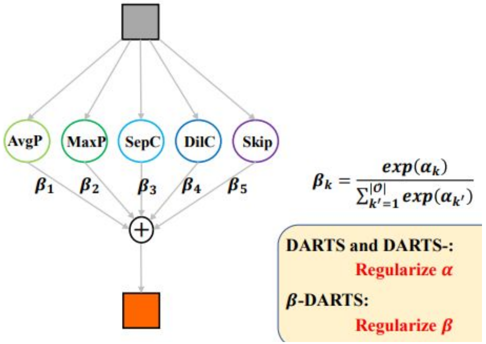
$$x^{(j)} = \sum_{i<j} o^{(i,j)}(x^{(i)})$$

$$\bar{o}^{(i,j)}(x) = \sum_{o \in \mathcal{O}} \frac{\exp(\alpha_o^{(i,j)})}{\sum_{o' \in \mathcal{O}} \exp(\alpha_{o'}^{(i,j)})} o(x)$$

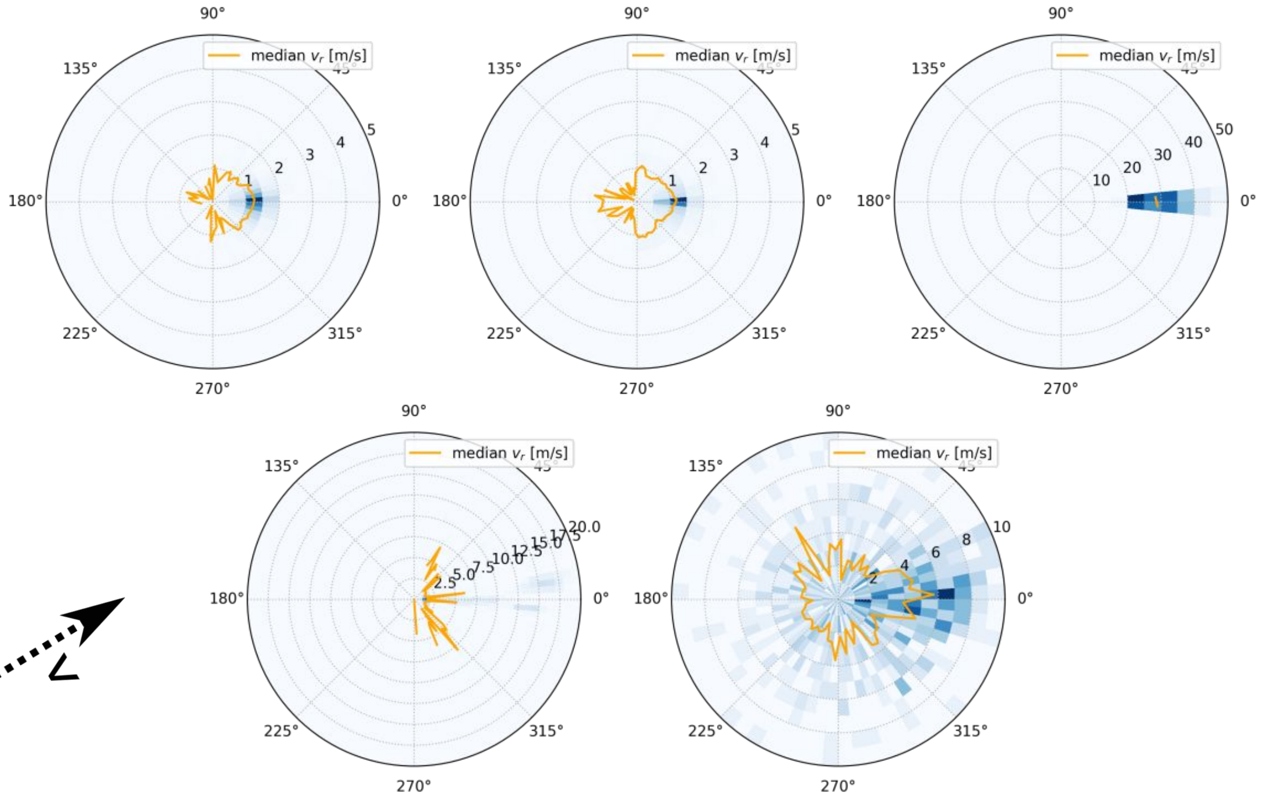
$$\min_{\alpha} \quad \mathcal{L}_{val}(w^*(\alpha), \alpha)$$
$$\text{s.t.} \quad w^*(\alpha) = \text{argmin}_w \ \mathcal{L}_{train}(w, \alpha)$$

Iteratively update architecture parameters on validation and network parameters on train set.
Gradient Based optimization. Weight sharing within supernet structure.

AML often surfers two main issues: weak robustness, poor generalization
To solve it, regularize DARTS. Beta-Decay regularization imposes constraints to keep the variance of activated architecture parameters small



Benchmarks


median v_r [m/s]
90°
135°
45°
180°
0°
225°
315°
270°
1
2
3
4
5
10
20
30
40
50
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
2
4
6
8
10