Waymo Motion Prediction Competition

3rd place solution

Our team

Artsiom Sanakoyeu
Heidelberg University

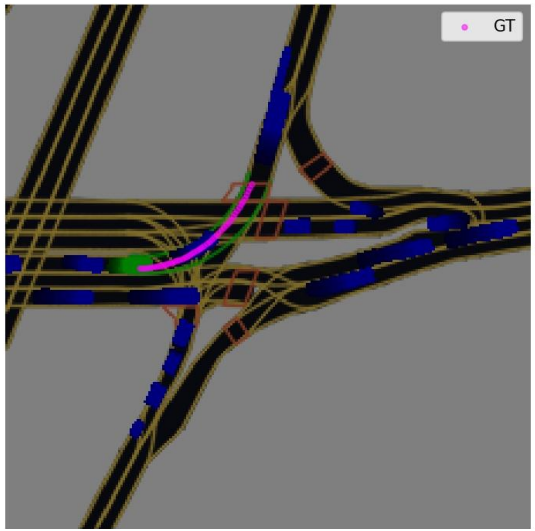
Kirill Brodt
Novosibirsk State University

Stepan Konev
Skoltech

Data

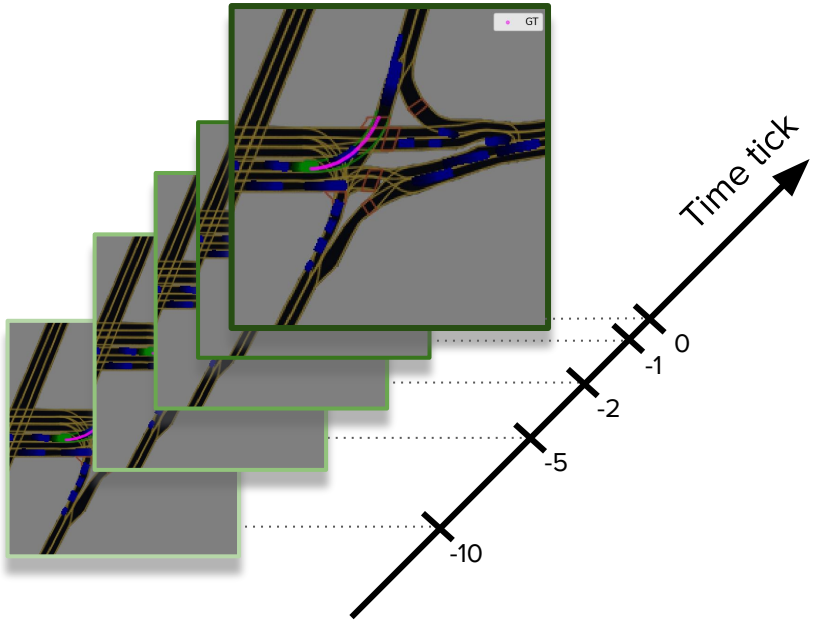
Goal: Predict future coordinates

Input image



224x224xC

History frames: current frame and 1 sec (10 frames) in the past

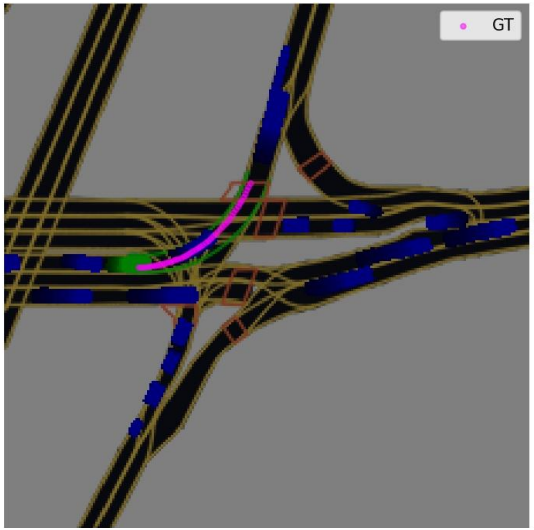


- Target agent's velocity at prediction moment is aligned with x-axis
- Target agent is always located in the same place at the image

Solution

Goal: Predict future coordinates

Input image



224x224xC

80 x 2d points and a scalar confidence

Trajectory 1

Trajectory 2

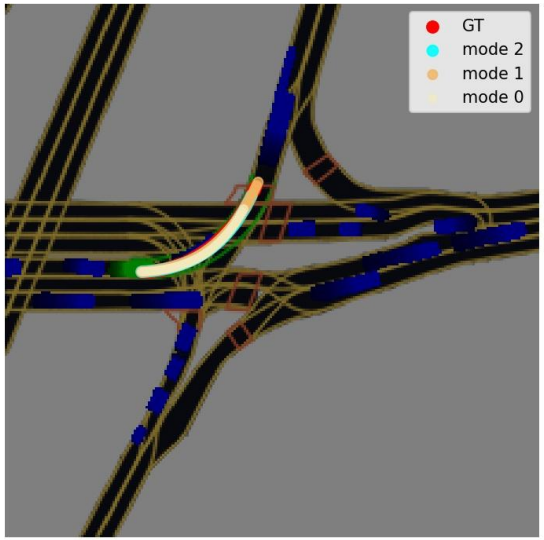
Trajectory 3

Trajectory 4

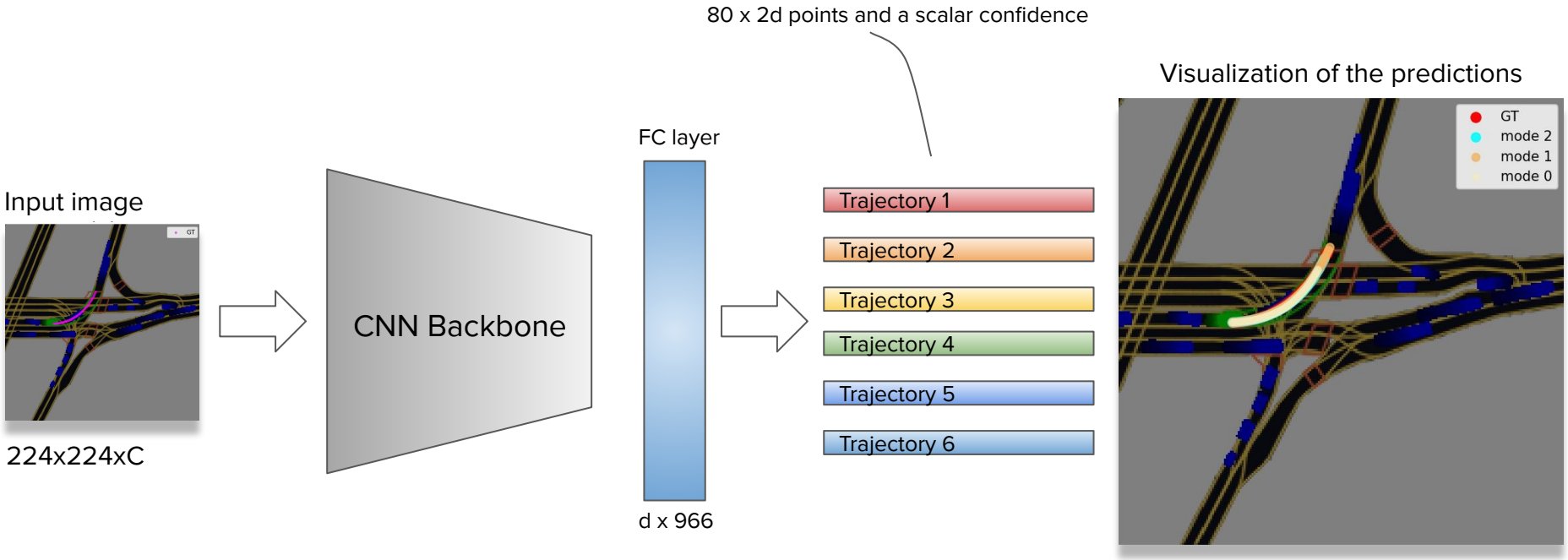
Trajectory 5

Trajectory 6

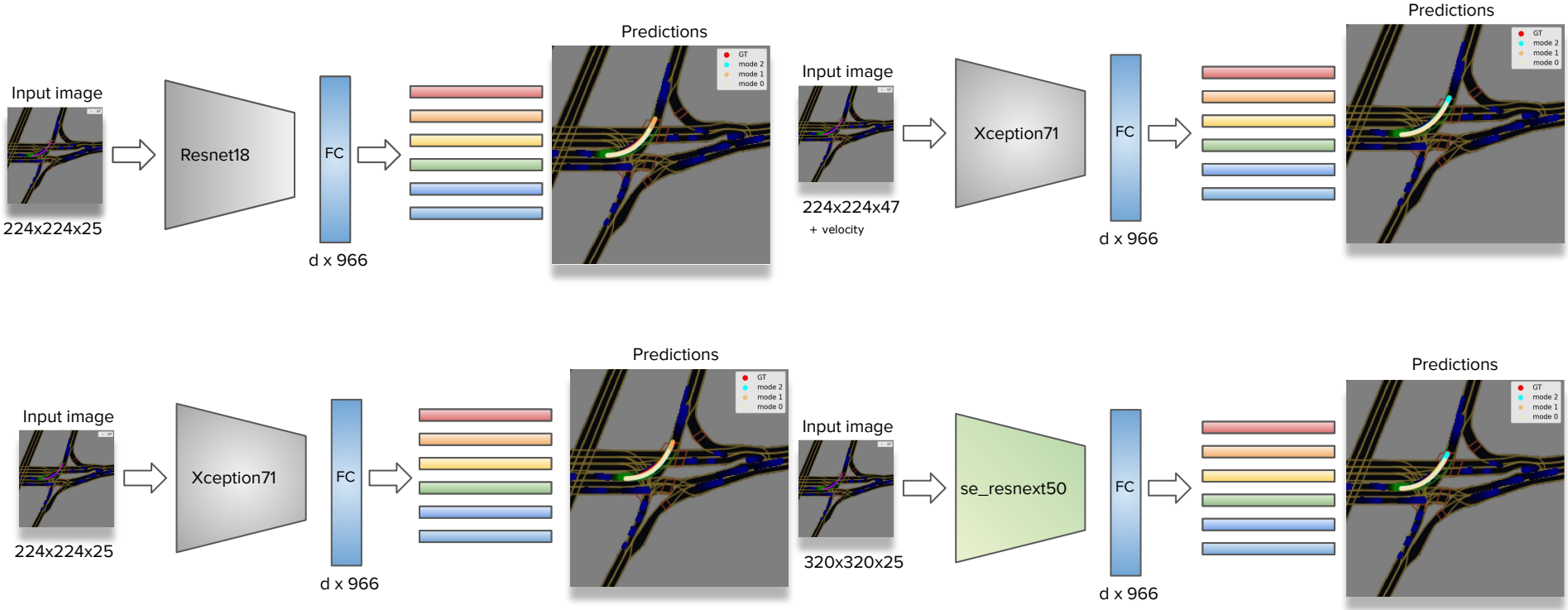
Visualization of the predictions



Core model: Regression CNN



Regression CNN: different backbones



Evaluation Metric and Loss

Ground truth coordinates

$$\mathbf{GT} = [\overbrace{(x_1, y_1), \ldots, (x_{80}, y_{80})}]$$

$$\mathbf{hypothesis}_k = [(x_1^{(k)}, y_1^{(k)}), \ldots, (x_{80}^{(k)}, y_{80}^{(k)})],\ k = 1, \ldots, K$$

$$L = -logP(\mathbf{GT}) =$$

Evaluation Metric and Loss

Ground truth coordinates

$$\mathbf{GT} = [\overbrace{(x_1, y_1), \ldots, (x_{80}, y_{80})}]$$

$$\mathbf{hypothesis}_k = [(x_1^{(k)}, y_1^{(k)}), \ldots, (x_{80}^{(k)}, y_{80}^{(k)})],\ k = 1, \ldots, K$$

$$L = -logP(\mathbf{GT}) =$$

Confidences — Predicted hypothesis #k

$$= -log \sum_k \overbrace{c^k} \prod_t \mathcal{N}(\mathbf{GT} | \overbrace{\mu = \mathbf{hypothesis}_k}, \Sigma = E)$$

Evaluation Metric and Loss

Ground truth coordinates

$$\mathbf{GT} = [\overbrace{(x_1, y_1), \ldots, (x_{80}, y_{80})}]$$

$$\mathbf{hypothesis}_k = [(x_1^{(k)}, y_1^{(k)}), \ldots, (x_{80}^{(k)}, y_{80}^{(k)})],\ k = 1, \ldots, K$$

$$L = -logP(\mathbf{GT}) =$$

Confidences

Predicted hypothesis #k

$$= -log \sum_k \overbrace{c^k} \prod_t \mathcal{N}(\mathbf{GT} | \overbrace{\mu = \mathbf{hypothesis}_k}, \Sigma = E)$$

$$= -log \sum_k c^k \prod_t \mathcal{N}(x_t | \mu = \bar{x}_t^{(k)}, \sigma = 1) \mathcal{N}(y_t | \mu = \bar{y}_t^{(k)}, \sigma = 1)$$

Evaluation Metric and Loss

Ground truth coordinates

$$\mathbf{GT} = [(x_1, y_1), \ldots, (x_{80}, y_{80})]$$

$$\mathbf{hypothesis}_k = [(x_1^{(k)}, y_1^{(k)}), \ldots, (x_{80}^{(k)}, y_{80}^{(k)})],\ k = 1, \ldots, K$$

$$L = -logP(\mathbf{GT}) =$$

Confidences — Predicted hypothesis #k

$$= -log \sum_k c^k \prod_t \mathcal{N}(\mathbf{GT} | \mu = \mathbf{hypothesis}_k, \Sigma = E)$$

$$= -log \sum_k c^k \prod_t \mathcal{N}(x_t | \mu = \bar{x}_t^{(k)}, \sigma = 1) \mathcal{N}(y_t | \mu = \bar{y}_t^{(k)}, \sigma = 1)$$

Confidences — Predicted coordinates — Ground truth coordinates

$$L = -\log \sum_k e^{\log(c^{(k)}) - \frac{1}{2} \sum_t (\bar{x}_t^{(k)} - x_t)^2 + (\bar{y}_t^{(k)} - y_t)^2}$$

Important training details: Rasterizer optimization

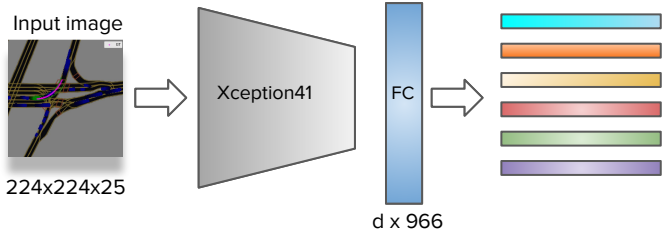
Default l5kit rasterizer: 15 images / sec on a single core.

Best single model: Xception71

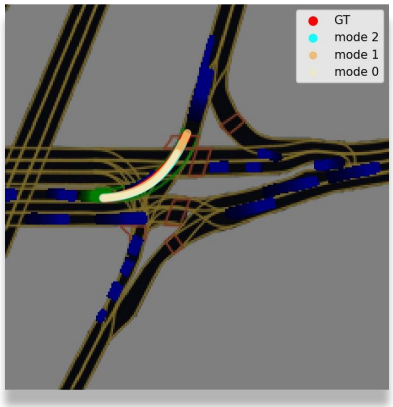
Test MAP score: **0.2136**

Training parameters:

- Adam with LR=0.001
- Batch size 48
- Scheduler: cosine annealing warm restarts every 2 epochs and minimum lr 1e-5



Top 3 Predicted hypotheses



Results

Object Type	**mAP**	Min ADE	Min FDE	Miss Rate	Overlap Rate
Vehicle	0.2357	0.8946	1.8175	0.2138	0.0886
Pedestrian	0.2175	0.4449	0.9131	0.1276	0.2725
Cyclist	0.1875	0.8803	1.7501	0.2860	0.1071
Avg	**0.2136**	0.7400	1.4936	0.2091	0.1560

Conclusion

Conclusion

- Rasterizer optimization is important
- A Simple CNN regression baseline is very strong
- Training for longer with right parameters on the full dataset was crucial
- Ensembling - most likely would give further improvements.

Regression CNN: More hypotheses

Confidences

Predicted coordinates

$$L = -\log \sum_k e^{\log(c^{(k)}) - \frac{1}{2}\sum_t (\bar{x}_t^{(k)} - x_t)^2 + (\bar{y}_t^{(k)} - y_t)^2}$$

Ground truth coordinates