Overcoming Limitations of Mixture Density Networks: A Sampling and Fitting Framework for Multimodal Future Prediction

YouTube Video

COMPUTER VISION University of Freiburg

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

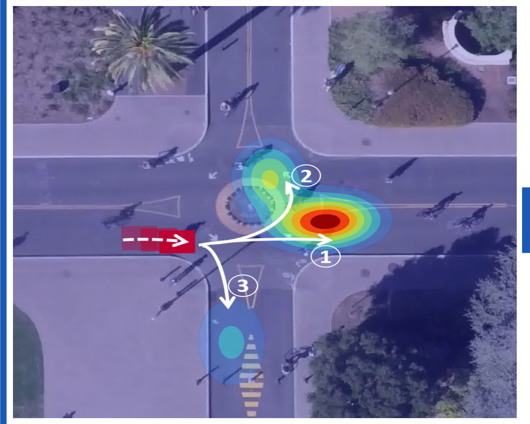
Osama Makansi makansio@cs.uni-freiburg.de

Eddy Ilg ilg@cs.uni-freiburg.de

Özgün Cicek cicek@cs.uni-freiburg.de

Thomas Brox brox@cs.uni-freiburg.de

Why do we need Multimodality & Uncertainty?



- The goals of agents are not observable.
- The system and its future are non-deterministic, thus **multimodal** and **uncertain**.

Challenges

- Environment constraints.
- Agents interactions.
- History Information.
- Statistical Information.

Past Methods & Ours

	Multimodality	Mixture Distribution	Diversity
Unimodal	✗	✗	✗
Sampling	✓	✗	✓
MDN [1]	✓	✓	✗
Ours	✓	✓	✓

MDNs predict general distributions but suffer from mode collapse.

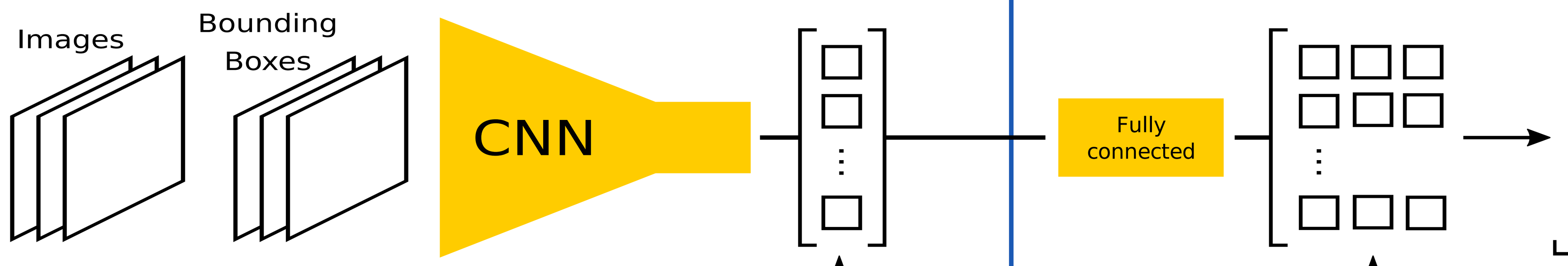
We address this by introducing a two-stages appraoch.

Watch video on YouTube: https://www.youtube.com/watch?v=bleGpgc2Odc

Our Contribution

Sampling

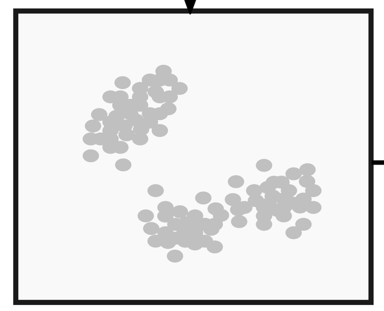
✓ Winner Takes All (WTA) ensures **diversity**



$$L_{WTA} = \sum_{k=1}^{K} w_k l(\boldsymbol{h}_k, \hat{\mathbf{y}})$$

- Vanilla WTA.
- Relaxed WTA.
- **Evolving WTA (Ours)**

$\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K$



Diverse Hypotheses

Fitting

✓ NLL provides unconstrained **multimodal distribution**

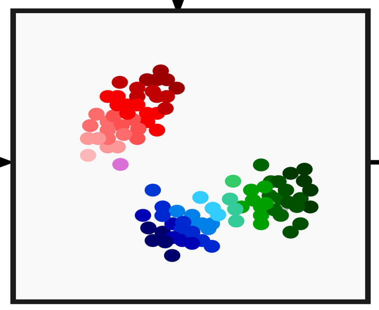
$$\gamma_k = \mathrm{softmax}(\boldsymbol{z}_k)$$

$$\pi_i = \frac{1}{K}\sum_{k=1}^{K}\gamma_{k,i} \qquad \boldsymbol{\mu}_i = \frac{\sum_{k=1}^{K}\gamma_{k,i}\boldsymbol{h}_k}{\sum_{k=1}^{K}\gamma_{k,i}}$$

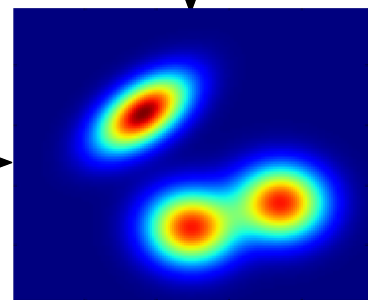
$$\boldsymbol{\sigma}_i^2 = \frac{\sum_{k=1}^{K}\gamma_{k,i}\left[(\boldsymbol{\mu}_i - \boldsymbol{h}_k)^2\right]}{\sum_{k=1}^{K}\gamma_{k,i}}$$

$$L_{NLL} = -\log\left[p(\mathbf{y}|\mathbf{x}) = \sum_{i=1}^{M}\pi_i \phi(\mathbf{y}|\boldsymbol{\theta}_i)\right]$$

$\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K$



Soft Assignments



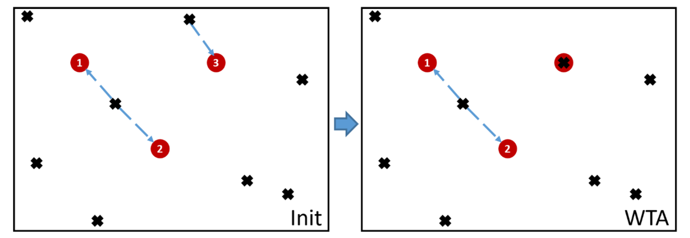
Mixture Distribution

Motivated by EM Algorithm

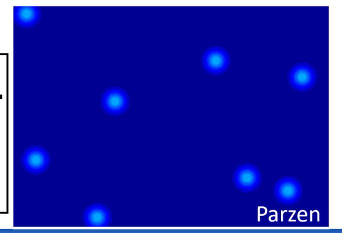
Multiple Hypotheses Prediction - Winner Takes All (WTA) Variants

WTA [4]

$$w_i = \delta(i = \underset{k}{\mathrm{argmin}}\, ||\boldsymbol{h}_k - \hat{\mathbf{y}}||)$$

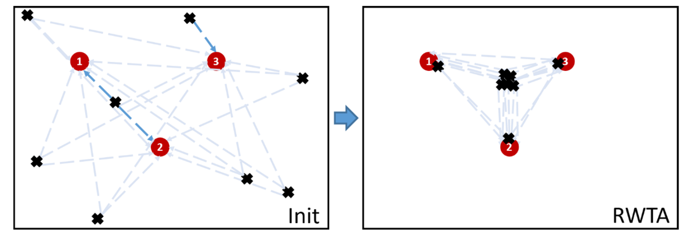


- Only one mode is matched.
- Many bad hypotheses.
>> wrong distribution.

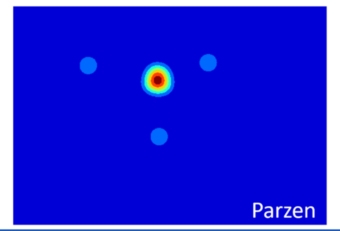


Relaxed WTA [2]

$$w_i = \begin{cases} 1-\epsilon & i = \underset{k}{\mathrm{argmin}}\, ||\boldsymbol{h}_k - \hat{\mathbf{y}}|| \\ \epsilon & else \end{cases}$$

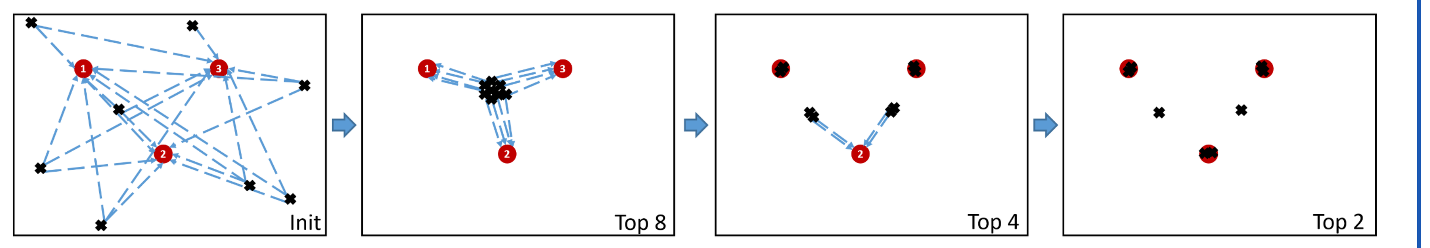


- All modes are matched.
- Many hypotheses to the center.
>> still wrong distribution.

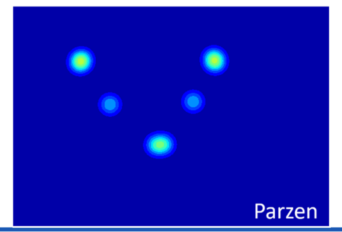


Evolving WTA (Ours)

while $N \neq 0$ **do**
$$w_i = \delta(i \in \underset{k_1,\ldots,k_N}{\mathrm{argmin}}\, ||\boldsymbol{h}_k - \hat{\mathbf{y}}||)\ ,\ N \leftarrow N/2$$

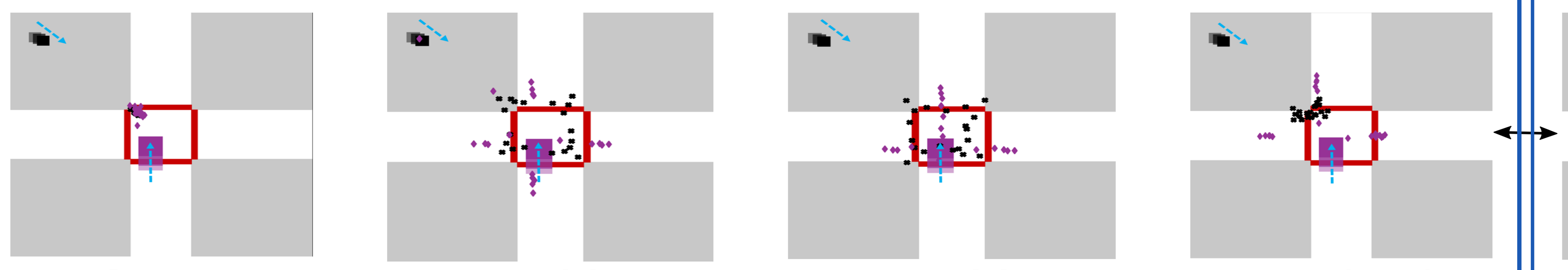


- All modes are matched
- Only few bad hypotheses are introduced.
>> closer to the true distribution



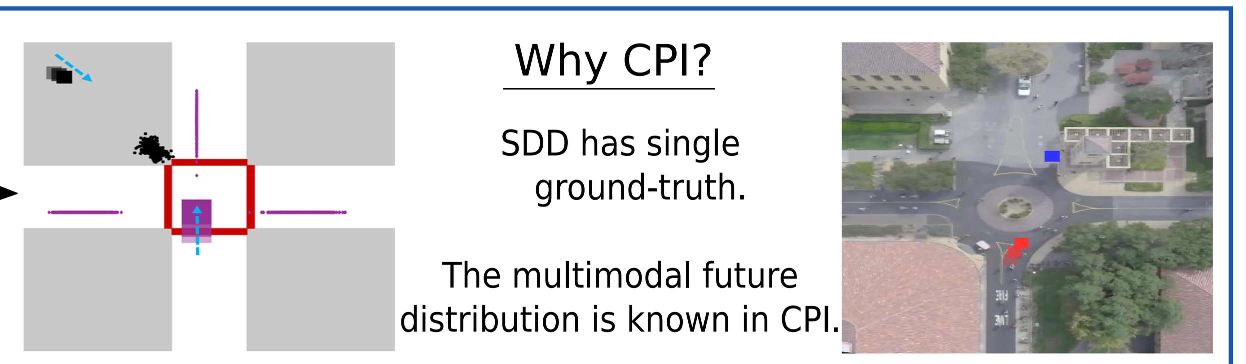
Results

Sampling on CPI



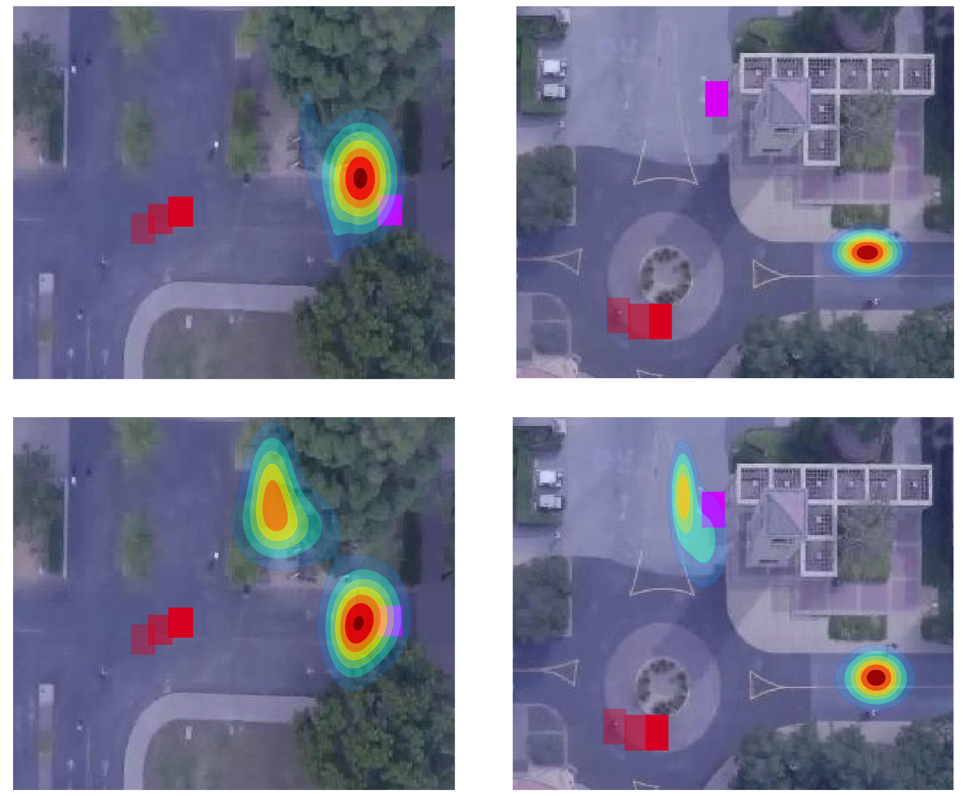
EWTA approximates the ground-truth distribution better.

Datasets

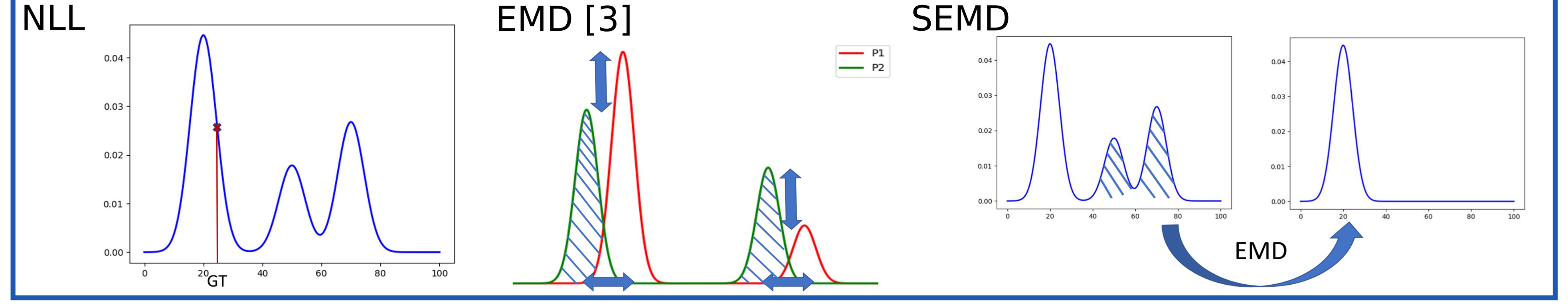


Car Pedestrian Interaction GT

Stanford Drone Dataset SDD [6] Real



Evaluation Metrics



Method	CPI NLL ↓	CPI EMD ↓	SDD NLL ↓	SDD SEMD ↑
Kalman Filter [5]	25.29	7.03	13.17	–
Unimodal Distribution	26.13	2.43	9.88	–
MDN [1]	9.20	1.83	9.71	2.36
Ours	**8.33**	**1.57**	**9.33**	**4.35**

- Ours outperforms all baselines on both synthetic data and real data.
- SEMD shows that ours enjoys more multimodality, while MDN suffers from mode collapse.
- Qualitative results support our findings.

References

[1] C. M Bishop. Mixture density networks. Technical report, Citeseer, 1994.
[2] C. Rupprecht, I. Laina, R. DiPietro, M. Baust, F. Tombari, N. Navab, G. DHager. Learning in an uncertain world: Representing ambiguity through multiple hypotheses (ICCV 2017).
[3] S. Shirdhonkar and D. W. Jacobs. Approximate earth movers distance in linear time (CVPR 2008).
[4] A. Rivera, D. Batra, and P. Kohli. Multiple choice learning: Learning to produce multiple structured outputs (NIPS 2012).
[5] R. E. Kalman. A new approach to linear filtering and prediction problems. ASME Journal of Basic Engineering, 1960.
[6] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese. Learning social etiquette: Human trajectory understanding in crowded scenes (ECCV 2016).

Acknowledgements

IMRA

Bundesministerium für Bildung und Forschung