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

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

Datasets
- Fully connected

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

References

Acknowledgements
- SEMD shows that ours enjoys more multimodality, while MDN suffers from mode collapse.
- Qualitative results support our findings.

Datasets
- Why CPI?
- SDD has single ground-truth.
- The multimodal future distribution is known in CPI.

Evaluation Metrics
- NLL
- EMD

Results
- Ours outperforms all baselines on both synthetic data and real data.
- Ours outperforms all baselines on both synthetic data and real data.
- Ours outperforms all baselines on both synthetic data and real data.

Our Contribution
- Winner Takes All (WTA) ensures diversity
- NLL provides unconstrained multimodal distribution
- Fitting

Motivated by EM Algorithm

Soft Assignments

Diverse Hypotheses

Images

Bounding Boxes

Sampling

Fully connected

Motivated by EM Algorithm

Sampling on CPI

EWTA (Ours)

Inputs

Outputs

Outputs

MDN (1)

Ours

Why CPI?
- SDD has single ground-truth.
- The multimodal future distribution is known in CPI.

Dropout
- MDN (1)
- Ours

WTA variants

- WTA [4]
- Relaxed WTA [2]
- Evolving WTA (Ours)
- Only one mode is matched. Many bad hypotheses. >> wrong distribution.
- All modes are matched. Many hypotheses to the center. >> still wrong distribution.
- All modes are matched. Only few bad hypotheses are introduced. >> closer to the true distribution

Why CPI?
- SDD has single ground-truth.
- The multimodal future distribution is known in CPI.

Results
- 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.