[o] Vision
COMPUTER VISION University of Freiburg
http://lmb.informatik.uni-freiburg.de

Uncertainty Estimates and Multi-Hypotheses Networks for Optical Flow

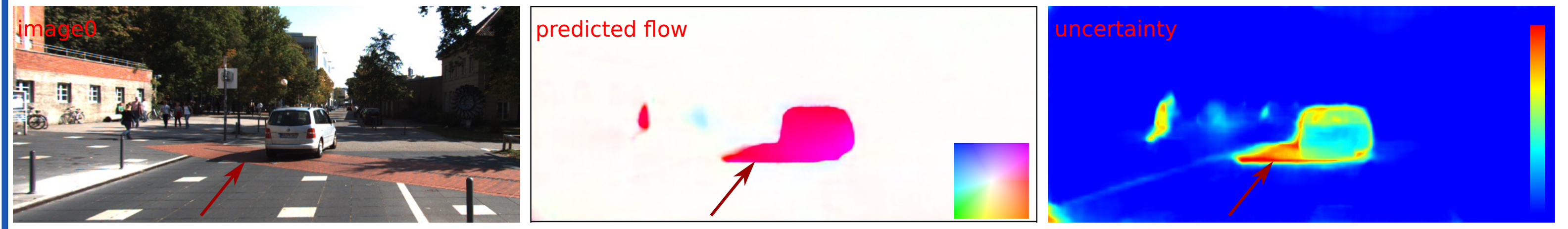
Eddy Ilg* ilg@cs.uni-freiburg.de
Özgün Çiçek* cicek@cs.uni-freiburg.de
Silvio Galesso* galessos@cs.uni-freiburg.de
Aaron Klein kleinaa@cs.uni-freiburg.de
Osama Makansi makansio@cs.uni-freiburg.de
Frank Hutter fh@cs.uni-freiburg.de
Thomas Brox brox@cs.uni-freiburg.de

* Equal Contribution

How Can We Trust CNNs?

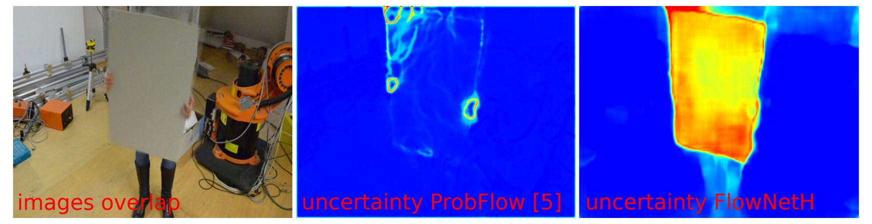
CNNs Today:
- Superior accuracy-runtime tradeoff
- Used in autonomous systems: self-driving cars
- Blackbox

CNNs We Want:
- Estimating also their reliability



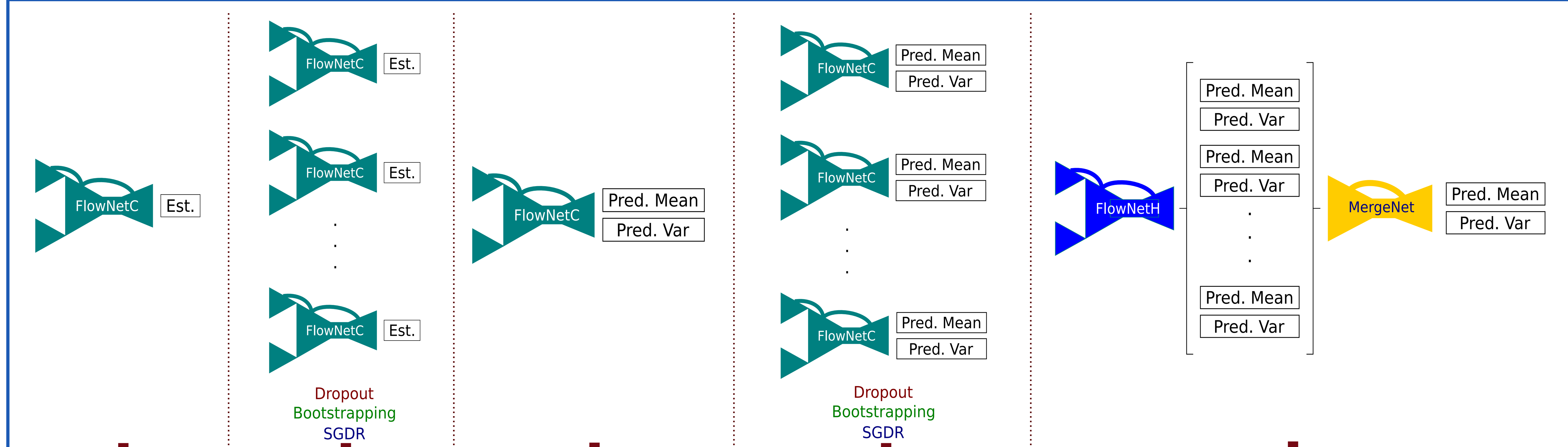
Our Contributions

- Investigating existing methods for uncertainty estimation in the scope of optical flow
  - MC Dropout [1]
  - Bootstrapped ensembles [2]
  - SGDR ensembles [3,4]
- Evaluating existing methods extensively:
  - Empirical
  - Predictive
  - Bayesian
- Proposing Multi-hypotheses network for state-of-the-art uncertainty estimation in realtime



Watch our video on YouTube: https://www.youtube.com/watch?v=HvyovWSo8uE

Overview of Investigated Approaches



Empirical

- Single FlowNetC [6,7]
- Cosine annealing for learning rate
- Training budget: 600k iterations
- Batch normalization
- Trained against EPE:

$$\mathrm{EPE} = \sqrt{(u-u^{gt})^2 + (v-v^{gt})^2}$$

Empirical Ensembles

Straightforward approach to get mean and variance from M individual networks:

$$\mu_{\mathbf{u}} = \frac{1}{M}\sum_{i=1}^{M} \mathbf{u}_{\mathbf{w}_i}(\mathbf{x})$$

$$\sigma_{\mathbf{u}}^2 = \frac{1}{M}\sum_{i=1}^{M} (\mathbf{u}_{\mathbf{w}_i}(\mathbf{x}) - \mu_{\mathbf{u}})^2$$

Predictive

Explicitely predicting mean and variance of a Laplace distribution by maximizing its log-likelihood:

$$\log p(\mathcal{D} \mid \mathbf{w}) = \frac{1}{N}\sum_{i=1}^{N} \log p(\mathbf{y}_i \mid \boldsymbol{\theta}(\mathbf{x}_i, \mathbf{w}))$$

$$\mathcal{L}(u|a,b) = \frac{1}{2b} e^{-\frac{|u-a|}{b}}$$

$$\mathcal{L}(u,v|a_u,a_v,b_u,b_v) \approx \mathcal{L}(u|a_u,b_u) \cdot \mathcal{L}(v|a_v,b_v)$$

$$-\log(\mathcal{L}(u|a_u,b_u) \cdot \mathcal{L}(v|a_v,b_v)) = \frac{|u-a_u|}{b_u} + \log b_u + \frac{|v-a_v|}{b_v} + \log b_v$$

Predictive Ensembles

Bayesian approach: ensembles of M predictive networks. The mean and the variance of the mixure distribution is computed by the law of total variance:

$$\boldsymbol{\mu}_{\mathbf{u}} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{\mu}_{\mathbf{u},i}$$

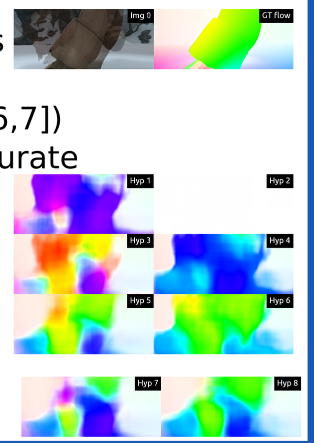
$$\boldsymbol{\sigma}_{\mathbf{u}}^2 = \frac{1}{M}\sum_{i=1}^{M} \left((\boldsymbol{\mu}_{\mathbf{u},i} - \boldsymbol{\mu}_{\mathbf{u}})^2 + \boldsymbol{\sigma}^2{}_{\mathbf{u},i}\right)$$

Multiple-Hypotheses Network

Desired:
- Obtain multiple samples in a single forward pass
- Multimodality in output hypothesis
- Merge M hypotheses with a network (FlowNetS[6,7])
- Diverse output to increase chances of being accurate

FlowNetH:
- Minimize:

$$L_{hyp} = \sum_{i,j} l(\boldsymbol{y}_{\text{best\_idx}(i,j)}, \boldsymbol{y}^{\text{gt}}(i,j)) + \Delta(i,j)$$



ECCV 2018 European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany

paper+poster:

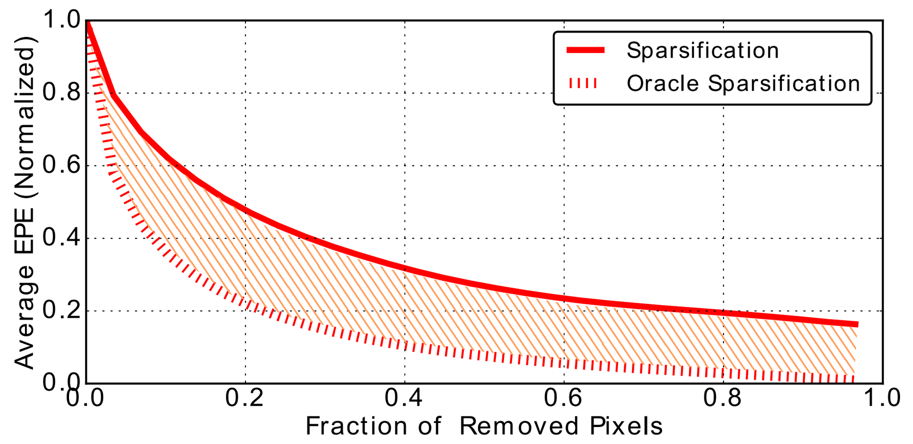
YouTube video:

Results

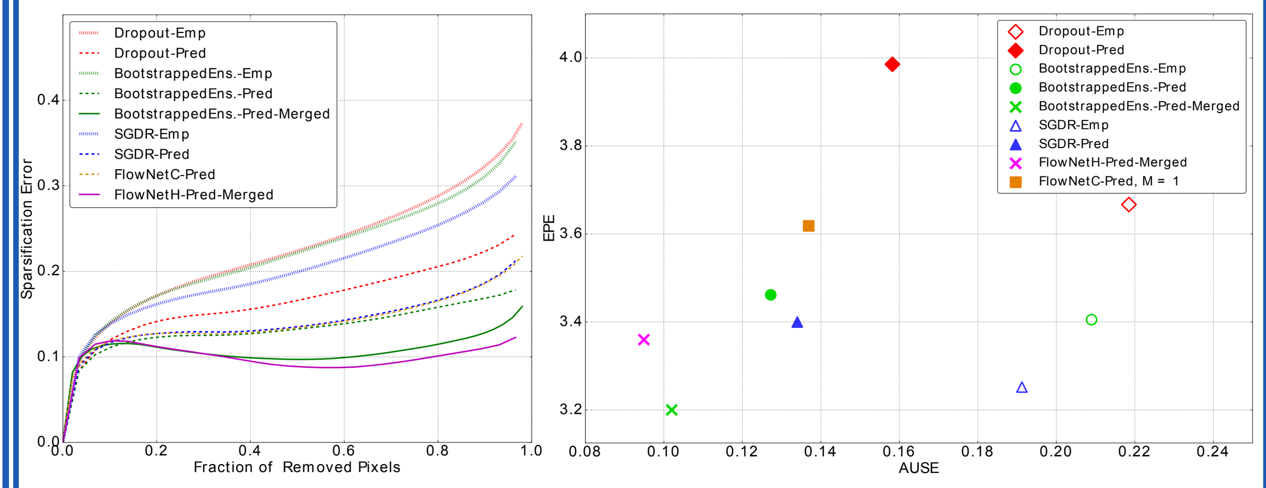
Evaluation Metrics

Sparsification: uncertainty based pixel ranking
Oracle Sparsification: EPE based pixel ranking

Our model independent metrics:
Sparsification Error: pointwise distance between curves
AUSE: Area Under Sparsification Error



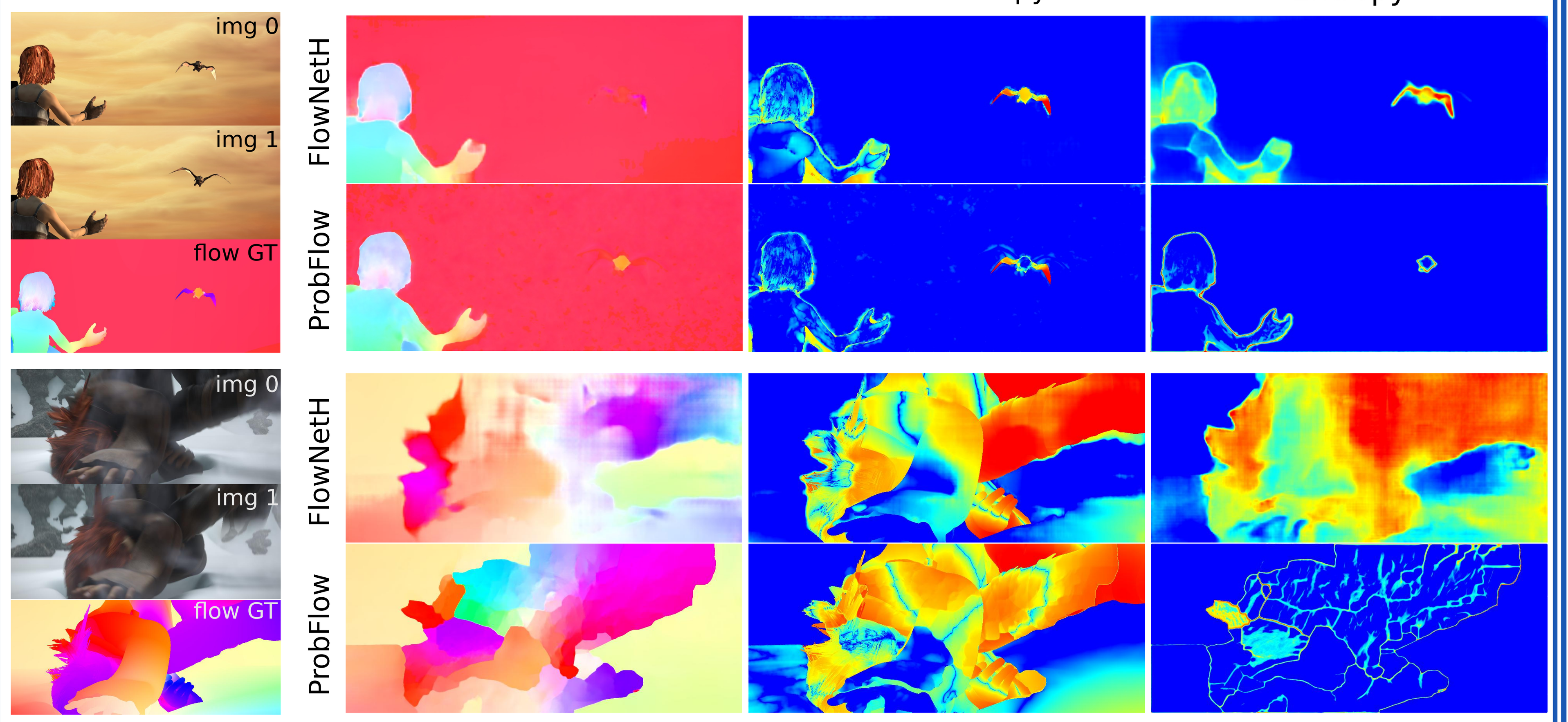
Method Comparison



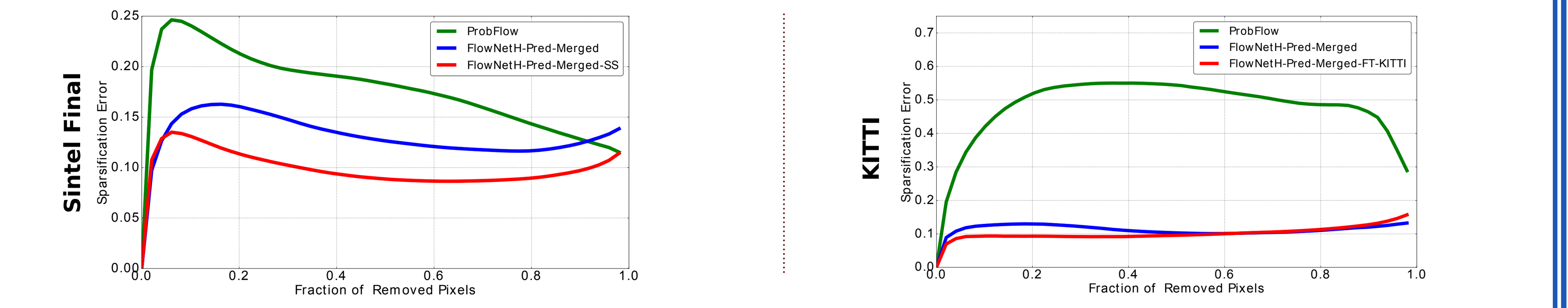
Results on Sintel Clean

	empirical (Emp)				predictive (Pred)				
	AUSE	EPE	Oracle EPE	Var.	AUSE	EPE	Oracle EPE	Var.	Runtime
FlowNetC	-	3.40	-	-	0.133	3.62	-	-	**38**ms
Dropout	0.212	3.67	2.56	5.05	0.158	3.99	2.96	3.80	320ms
SGDREnsemble	**0.191**	3.25	2.56	3.50	0.134	3.40	2.87	1.52	304ms
BootstrappedEnsemble	0.209	3.41	2.17	9.52	0.127	3.46	2.49	6.15	304ms
BootstrappedEnsemble-Merged					0.102	3.20	2.49	6.15	332ms
FlowNetH-Merged	-	3.50	1.73	**83.32**	**0.095**	3.36	1.89	**52.85**	60ms

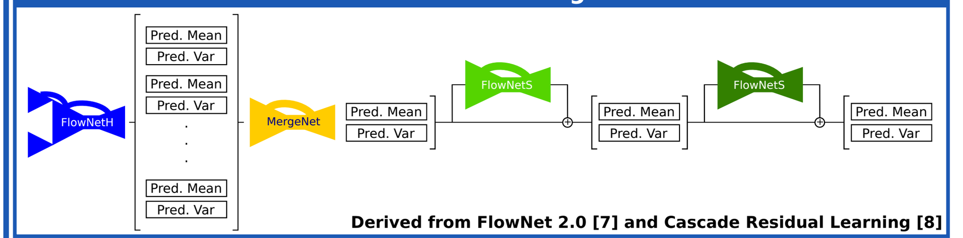
Comparison to the State of the Art



Sparsification Error Curves:



FlowneetH-Pred-Merged-SS



Stacked Networks and Comparison to ProbFlow [5]:

	Sintel Clean		Sintel Final		KITTI		runtime
	AUSE	EPE	AUSE	EPE	AUSE	EPE	
ProbFlow	0.162	1.87	0.173	3.34	0.466	8.95	38.1s†
FlowNetH-Pred-Merged-FT-KITTI	-	-	-	-	**0.086**	3.12	**60**ms
FlowNetH-Pred-Merged	0.117	2.58	0.128	3.78	0.151	7.84	**60**ms
FlowNetH-Pred-Merged-S	0.091	2.29	0.098	3.51	0.102	6.86	86ms
FlowNetH-Pred-Merged-SS	**0.089**	2.19	**0.096**	3.40	0.091	6.50	99ms

Conclusions

Emp. Ensembles vs. Pred. Ensembles:
- empirical ensembles provide more diversity
- empirical ensembles have lower EPE
- predictive ensembles have better AUSE
- predictive ensembles have no big advantage over a single predictive model

Ensemble Types:
- **Dropout**: worst AUSE and EPE
- **SGDR**: best EPE and good AUSE
- **BootstrappedEns**: best AUSE

Merging Networks: best among all internal methods
- Models with merging network are on-par on EPE and AUSE
- **FlowNetH** has the best accuracy-runtime tradeoff

FlowNetH-Pred-Merged-SS vs. ProbFlow [5]:

Overall, FlowNetH-Pred-Merged-SS is the best in terms of AUSE, EPE and runtime except EPE on Sintel, where it is slightly worse than ProbFlow.

References

[1] Kendall, A., Gal, Y.: What Uncertainties Do We Need in Bayesian Deep Learn-ing for Computer Vision? In: NIPS. (2017)
[2] Lakshminarayanan, B., Pritzel, A., Blundell, C.: Simple and scalable predictiveuncertainty estimation using deep ensembles. In: NIPS workshop. (2016)
[3] Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts.In: ICLR. (2017)
[4] Huang, G., Li, Y., Pleiss, G.: Snapshot ensembles: Train 1, get M for free. In: ICLR. (2017)
[5] Wannenwetsch, A.S., Keuper, M., Roth, S.: Probflow: Joint optical flow and un-certainty estimation. In: ICCV. (Oct2017)
[6] Dosovitskiy, A., Fischer, P., Ilg, E., H¨ausser, P., Hazırba¸s, C., Golkov, V., v.d.Smagt, P., Cremers, D., Brox, T.: Flownet: Learning optical flow with convolutional networks. In: ICCV. (2015)
[7] Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T.: Flownet 2.0:Evolution of optical flow estimation with deep networks. In: CVPR. (2017)
[8] Pang, J., Sun, W., Ren, J.S.J., Yang, C., Yan, Q.: Cascade residual learning: A two-stage convolutional neural network for stereo matching. In: ICCV Workshop. (2017)

Acknowledgements

We gratefully acknowledge funding by the German Research Foundation (SPP 1527 grants BR 3815/8-1 and HU 1900/3-1, CRC-1140 KIDGEM Z02) and by the Horizon 2020 program of the EU via the project Trimbot2020.