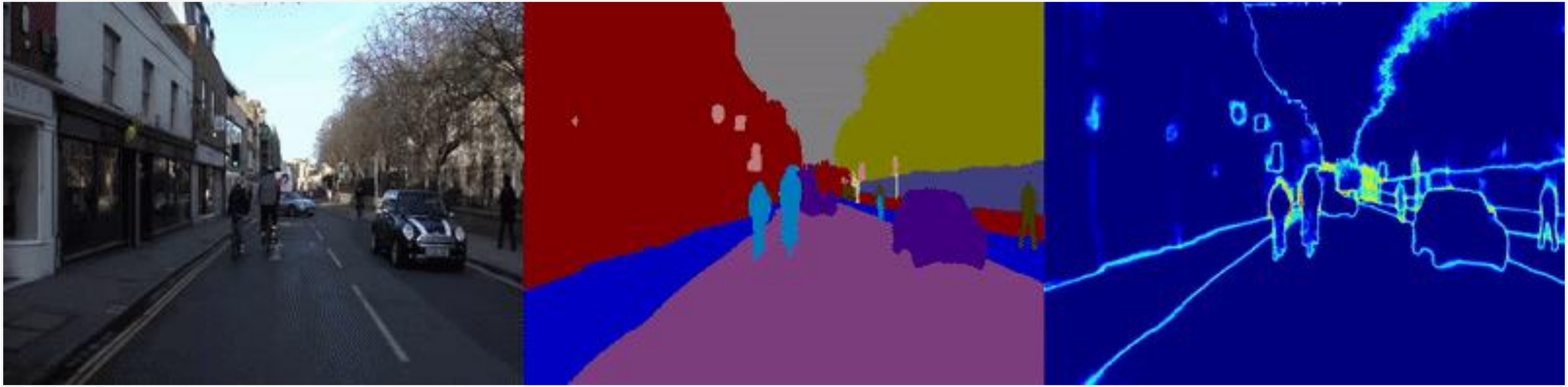
Geometry and Uncertainty in Deep Learning for Computer Vision

Alex Kendall, University of Cambridge, March 2017

@alexgkendall alexgkendall.com agk34@cam.ac.uk

Why is uncertainty important?

Bayesian SegNet for probabilistic scene understanding



Input Image | Semantic Segmentation | Uncertainty

Outline of Talk

1. What **uncertainty** can we model with deep learning and what are the benefits?
2. How do we model uncertainty using **Bayesian deep learning** for regression and classification tasks?
3. Why should we formulate deep learning models for vision which leverage our knowledge of **geometry**?

Uncertainty

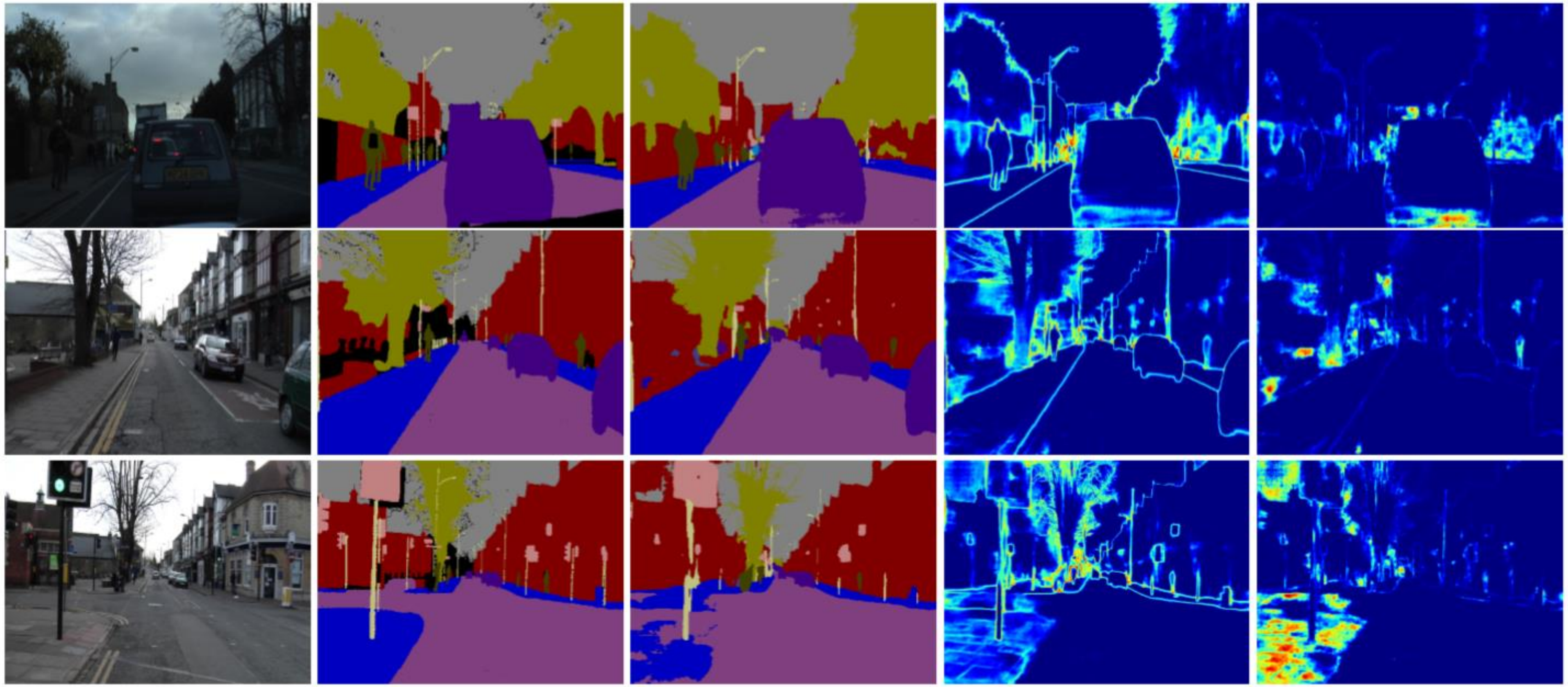
What kind of uncertainty can we model?

1. *Epistemic* uncertainty
   - Measures what you're model doesn't know
   - Can be explained away by unlimited data
2. *Aleatoric* uncertainty
   - Measures what you can't understand from the data
   - Can be explained away by unlimited sensing

Alex Kendall and Yarin Gal. **What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?** arXiv preprint 1703.04977, 2017.

What kind of uncertainty can we model?

Epistemic uncertainty is *modeling* uncertainty
Aleatoric uncertainty is *sensing* uncertainty



(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

Modeling Uncertainty with Bayesian Deep Learning

?

Deep learning is required to achieve state of the art results in computer vision applications but doesn't provide uncertainty estimates.

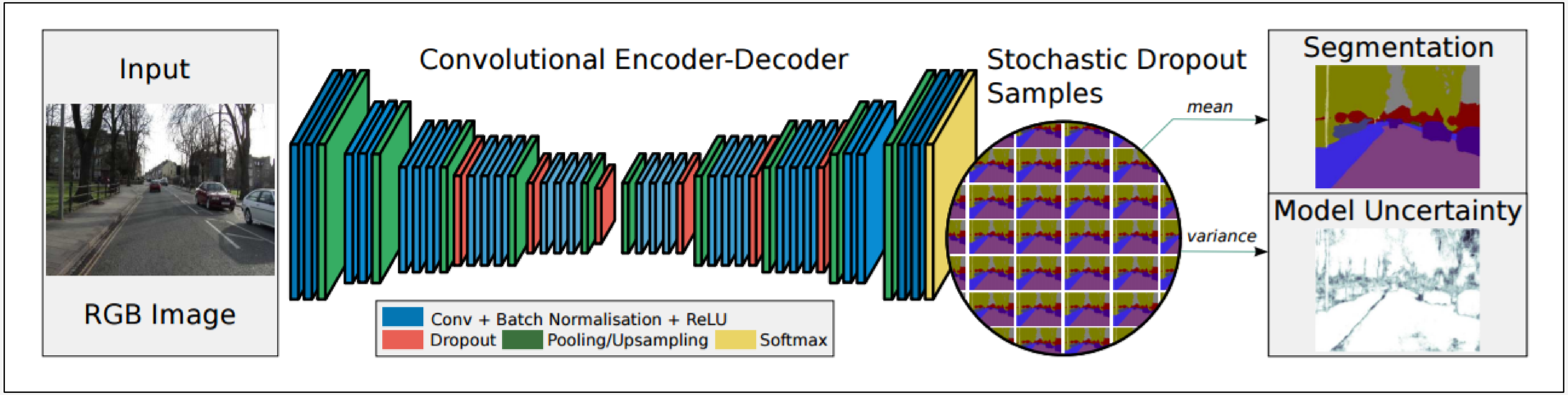
- **Bayesian neural networks** are a framework for understanding uncertainty in deep learning
- They have **distributions over network parameters** (rather than deterministic weights)
- Traditionally they have been **tricky to scale**

MacKay, David JC. **A practical Bayesian framework for backpropagation networks.** Neural Computation, 4(3): 448–472, 1992.

Modeling Epistemic Uncertainty with Bayesian Deep Learning

We can **model epistemic uncertainty** in deep learning models using Monte Carlo **dropout sampling** at test time.

Dropout sampling can be interpreted as **sampling from a distribution over models**.



Alex Kendall, Vijay Badrinarayanan and Roberto Cipolla **Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding**. arXiv preprint arXiv:1511.02680, 2015.

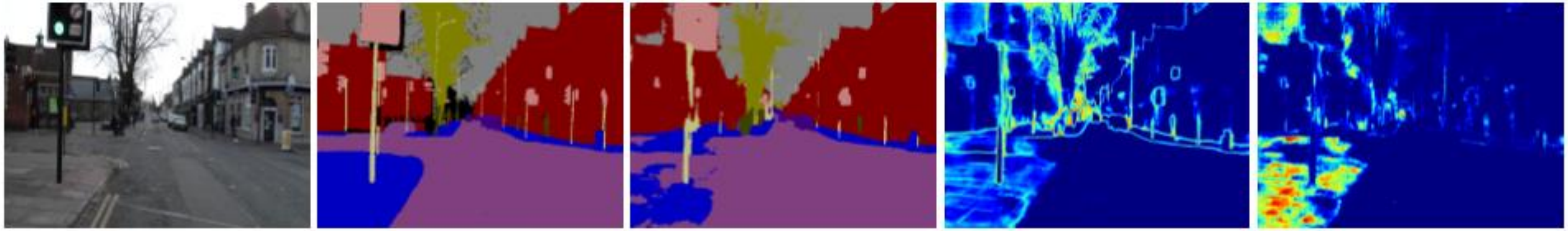
Modeling Aleatoric Uncertainty with Probabilistic Deep Learning

	Deep Learning	Probabilistic Deep Learning
Model	$[\hat{y}] = f(x)$	$[\hat{y}, \hat{\sigma}^2] = f(x)$
Regression	$Loss = \|y - \hat{y}\|^2$	$Loss = \frac{\|y - \hat{y}\|^2}{2\hat{\sigma}^2} + \log \hat{\sigma}^2$
Classification	$Loss = SoftmaxCrossEntropy(\hat{y}_{\mathrm{t}})$	$\hat{y}_{\mathrm{t}} = \hat{y} + \epsilon_t \qquad \epsilon_t \sim N(0, \hat{\sigma}^2)$ $Loss = \frac{1}{T}\sum_t SoftmaxCrossEntropy(\hat{y}_{\mathrm{t}})$

Alex Kendall and Yarin Gal. **What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?** arXiv preprint 1703.04977, 2017.

Semantic Segmentation Performance on CamVid

CamVid Results	IoU Accuracy
DenseNet (State of the art baseline)	67.1
+ Aleatoric Uncertainty	67.4
+ Epistemic Uncertainty	67.2
+ Aleatoric & Epistemic	**67.5**

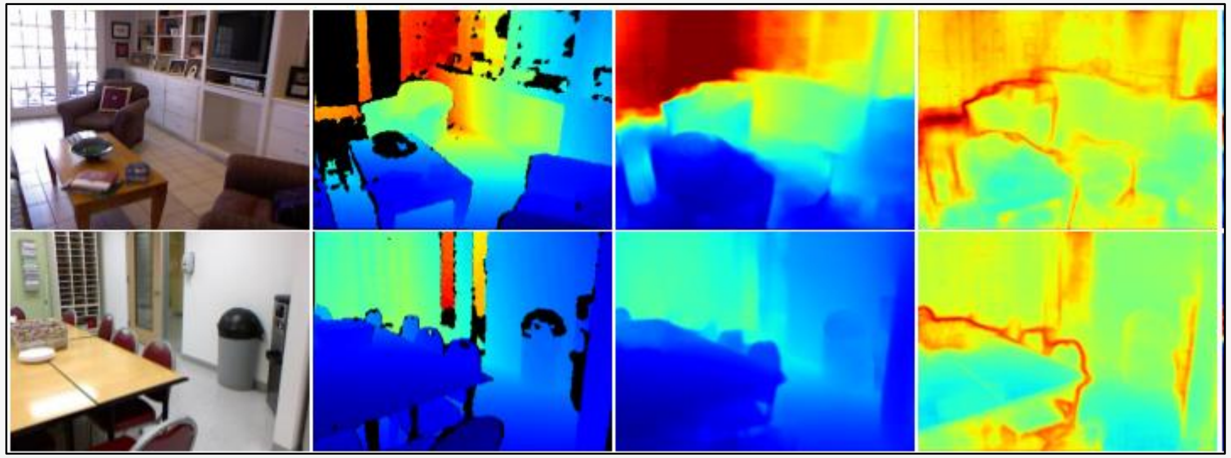


(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

Alex Kendall and Yarin Gal. **What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?** arXiv preprint 1703.04977, 2017.

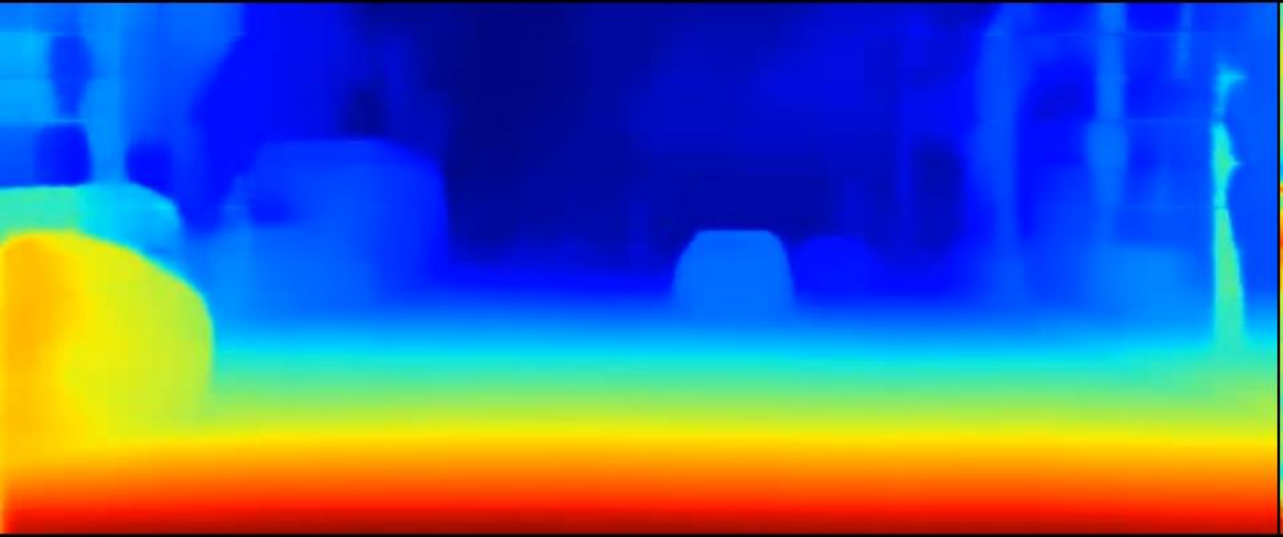
Monocular Depth Regression Performance

NYU Depth Results	Rel. Error
DenseNet (State of the art baseline)	0.167
+ Aleatoric Uncertainty	0.149
+ Epistemic Uncertainty	0.162
+ Aleatoric & Epistemic	**0.145**

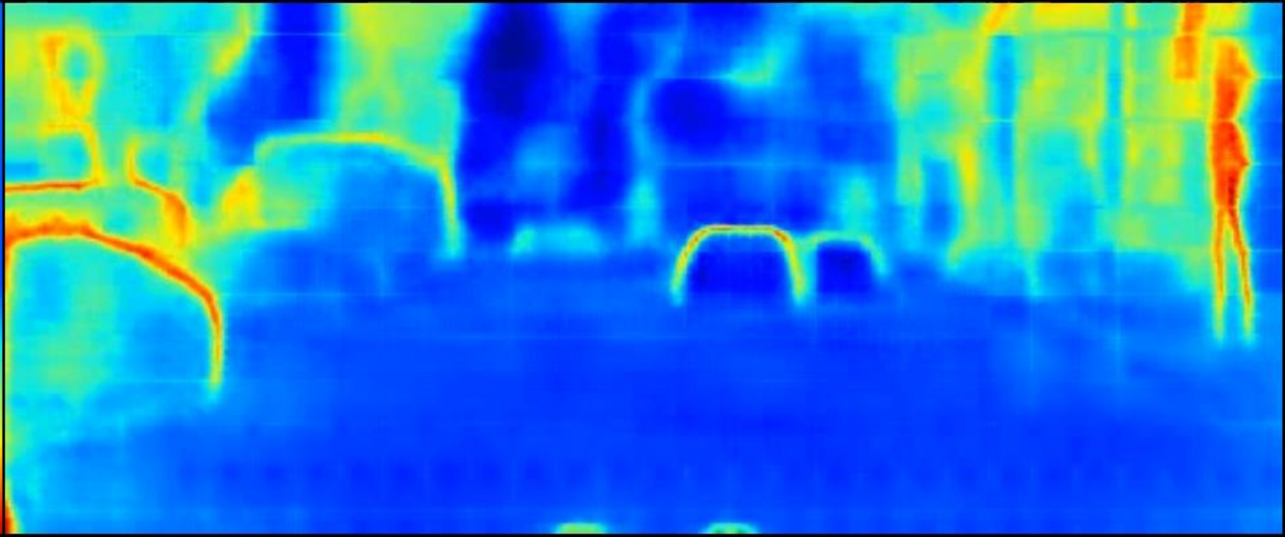


Alex Kendall and Yarin Gal. **What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?** arXiv preprint 1703.04977, 2017.

Input Video (Monocular)



Predicted Depth



Uncertainty

Aleatoric vs. Epistemic Uncertainty for Out of Dataset Examples

Train dataset	Test dataset	RMS	Aleatoric variance	Epistemic variance
Make3D / 4	Make3D	5.76	0.506	7.73
Make3D / 2	Make3D	4.62	0.521	4.38
Make3D	Make3D	3.87	0.485	2.78
Make3D / 4	NYUv2	-	0.388	15.0
Make3D	NYUv2	-	0.461	4.87

!

Aleatoric uncertainty remains constant while epistemic uncertainty increases for out of dataset examples!

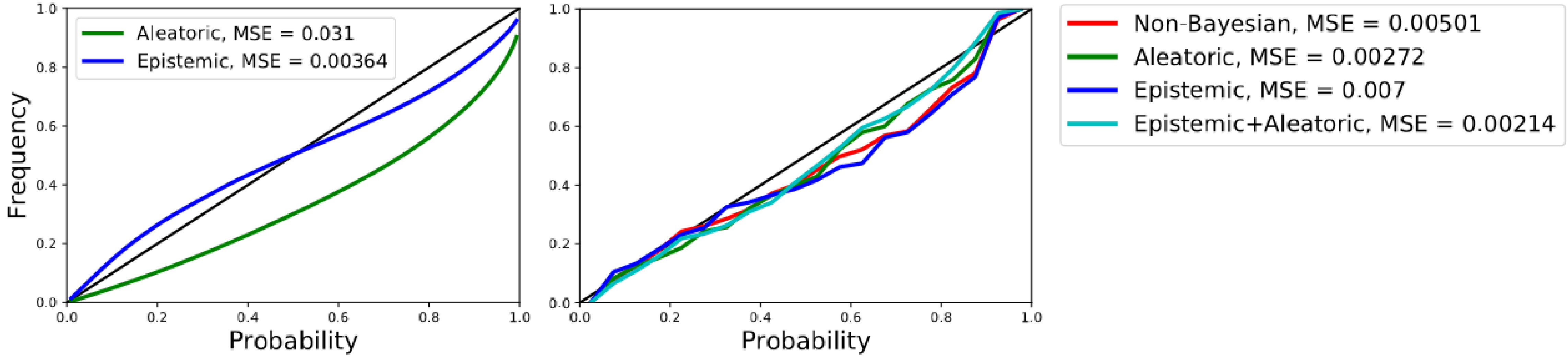
Alex Kendall and Yarin Gal. **What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?** arXiv preprint 1703.04977, 2017.

Uncertainty Benchmarks

- One reason why computer vision has progressed so rapidly is because we can benchmark and compare algorithms easily
- Often leaderboards rank prediction accuracy and algorithm speed
- *Leaderboards should also rank algorithms probabilistically and quantify uncertainty accuracy*

Calibration Plots

- For a prediction with probability p, the model should be correct with a frequency of p
- Perfect calibration corresponds to the line, $y = x$

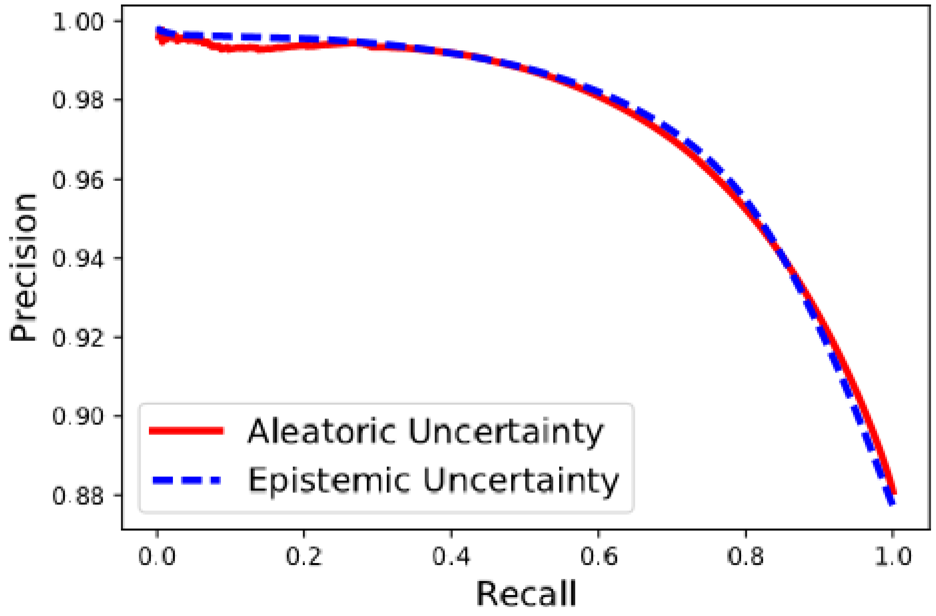


(a) Regression (Make3D) (b) Classification (CamVid)

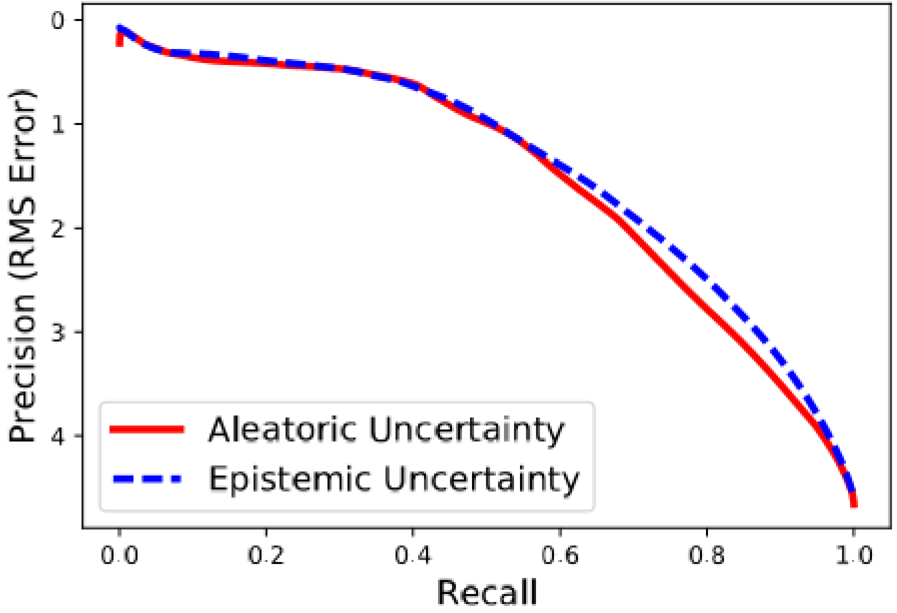
Alex Kendall and Yarin Gal. **What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?** arXiv preprint 1703.04977, 2017.

Precision Recall Plots

- Uncertainty should correlate well with accuracy



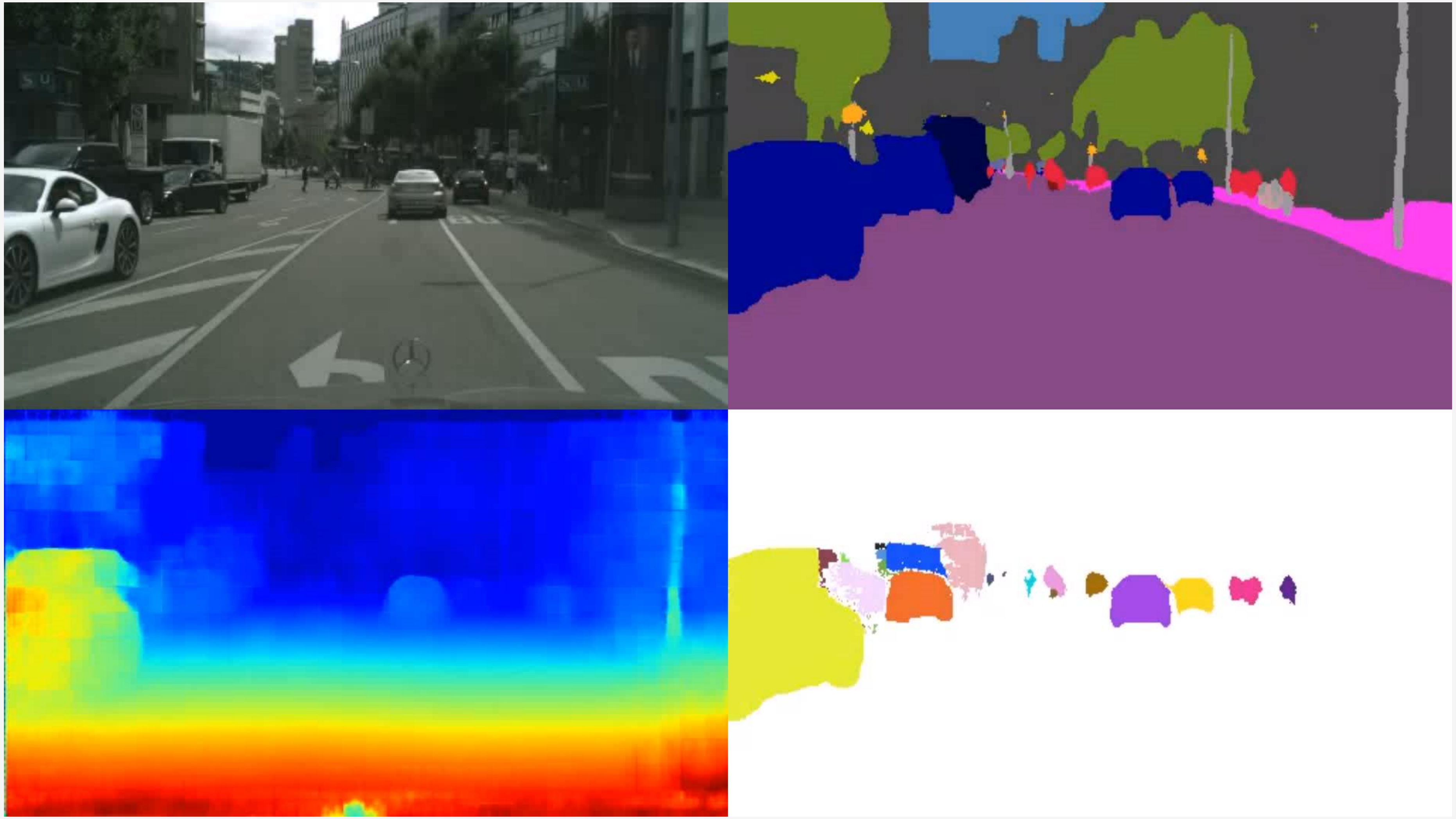
(a) Classification (CamVid)



(b) Regression (Make3D)

Alex Kendall and Yarin Gal. **What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?** arXiv preprint 1703.04977, 2017.

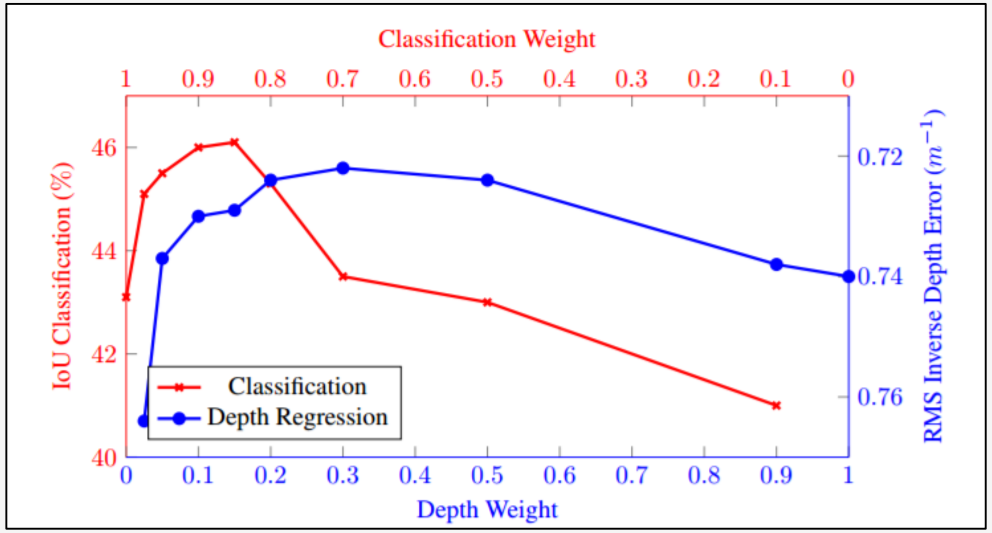
Putting it all Together: Multi-Task Learning



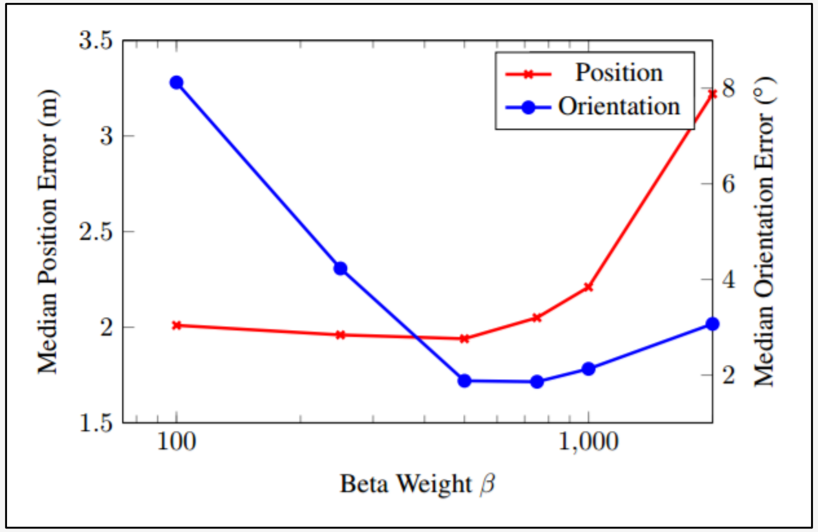
Multitask Learning

We want to simultaneously learn multiple tasks: $Loss = \sum_i w_i L_i$

Task performance is very sensitive to choice of weights, so how do you choose??



Scene Understanding [1]



Localisation [2]

[1] Alex Kendall, Yarin Gal and Roberto Cipolla. **Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics.** arxiv preprint 1705.07115, 2017.
[2] Alex Kendall, Matthew Grimes and Roberto Cipolla **PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization.** ICCV, 2015.

Types of Aleatoric Uncertainty

1. *Heteroscedastic* aleatoric uncertainty
   - Data dependent aleatoric uncertainty

2. *Homoscedastic* aleatoric uncertainty
   - Aleatoric uncertainty which doesn’t depend on the data
   - Task uncertainty

Alex Kendall and Yarin Gal. **What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?** arXiv preprint 1703.04977, 2017.

Combine Losses Using Homoscedastic Uncertainty

Homoscedastic uncertainty, σ^2, captures uncertainty of the entire task itself – not dependant on input data.

We propose to use this to learn a weighting for each loss term.

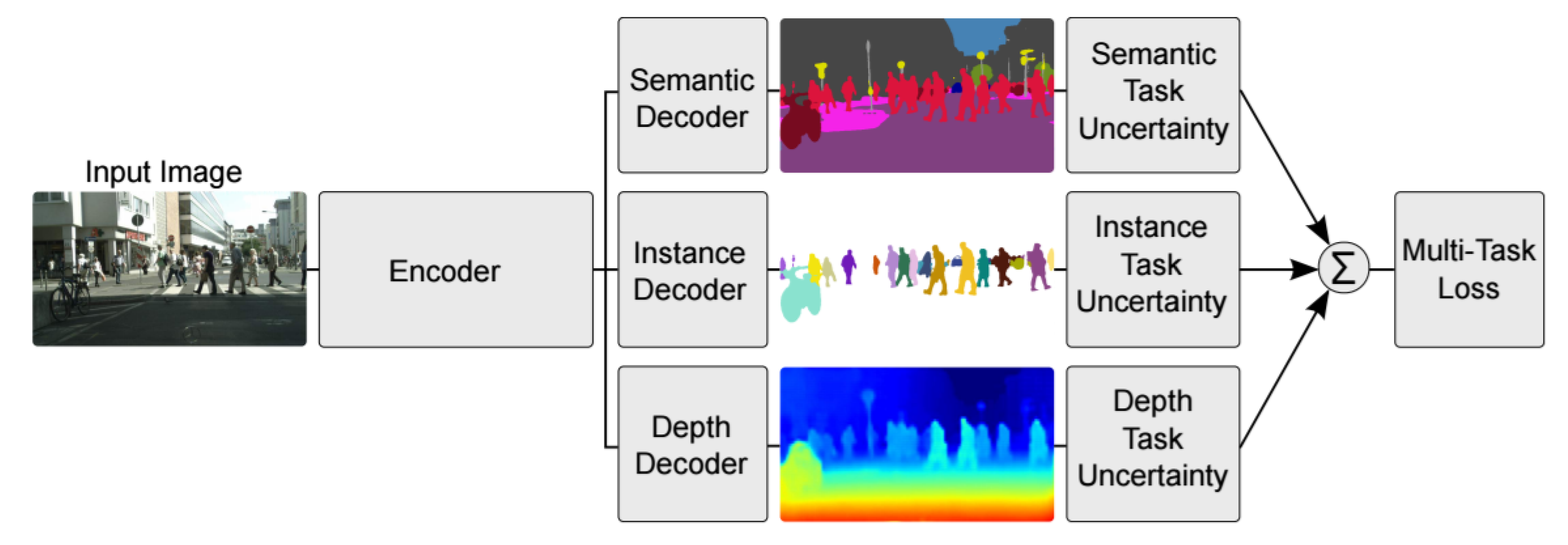
$$Loss = \overbrace{\frac{L_{regression\,1}}{\sigma_1^2} + \log \sigma_1^2}^{\text{Depth Regression}} + \overbrace{\frac{L_{regression\,2}}{\sigma_2^2} + \log \sigma_2^2}^{\text{Instance Regression}} + \overbrace{SoftmaxCrossEntropy\left(\frac{y}{\sigma_3^2}\right)}^{\text{Semantic Segmentation}}$$

Learnt uncertainty tempers each loss term

Regularizes homoscedastic uncertainty from going to infinity

[1] Alex Kendall, Yarin Gal and Roberto Cipolla. **Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics.** arxiv preprint 1705.07115, 2017.

Multi Task Scene Understanding Model



[1] Alex Kendall, Yarin Gal and Roberto Cipolla. **Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics.** arxiv preprint 1705.07115, 2017.

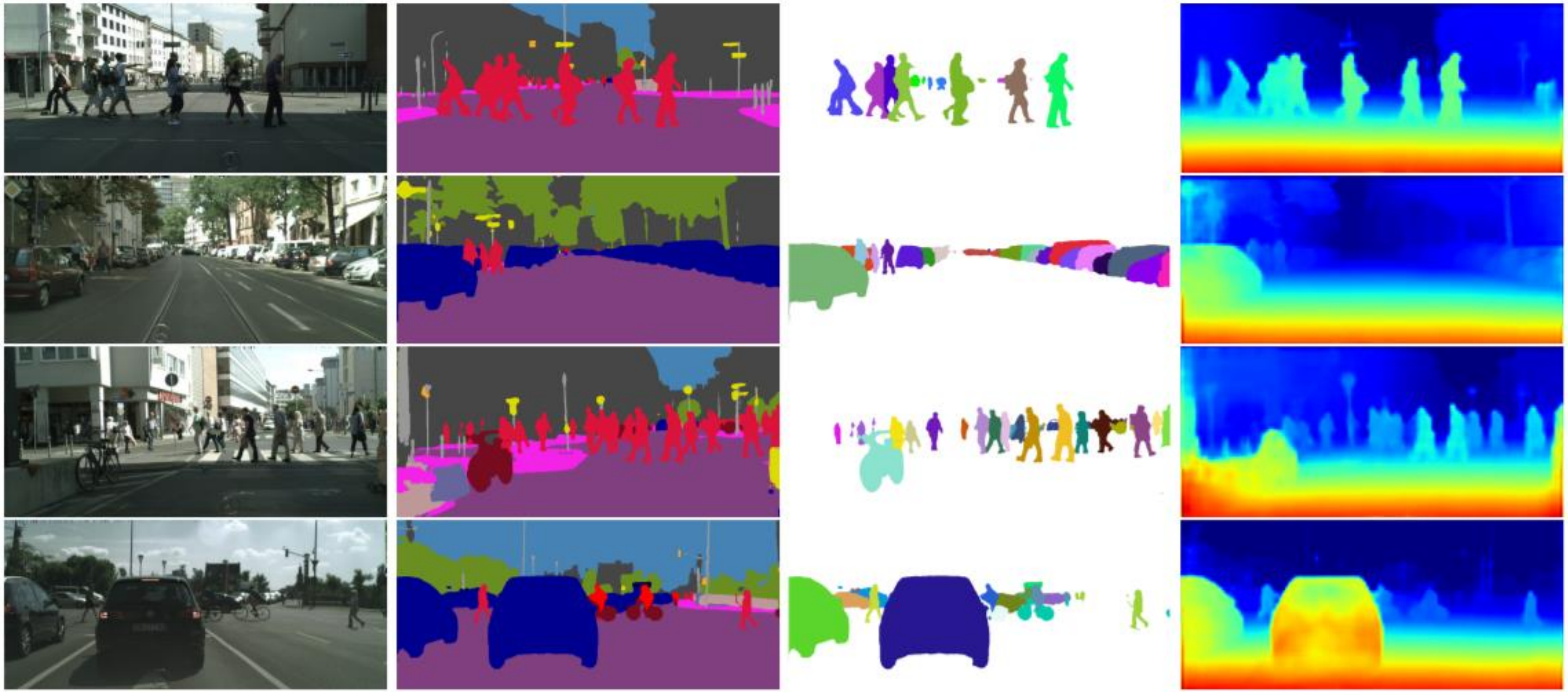
Multitask Learning Results

- Homoscedastic uncertainty can learn the optimal weighting
- Multitask learning can improve performance compared with training separate models for each individual task

Loss	Task Weights Cls.	Inst.	Depth	Classification IoU [%]	Instance RMS Error [px]	Inverse Depth RMS Error [px]
Class only	1	0	0	43.1%	-	-
Instance only	0	1	0	-	4.61	-
Depth only	0	0	1	-	-	0.783
Unweighted sum of losses	0.333	0.333	0.333	43.6%	3.92	0.786
Approx. optimal weights	0.8	0.05	0.15	46.3%	3.92	0.799
2 task uncertainty weighting	✓	✓		46.5%	3.73	-
2 task uncertainty weighting	✓		✓	46.2%	-	0.714
2 task uncertainty weighting		✓	✓	-	4.06	0.744
3 task uncertainty weighting	✓	✓	✓	**46.6%**	**3.91**	**0.702**

[1] Alex Kendall, Yarin Gal and Roberto Cipolla. **Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics.** arxiv preprint 1705.07115, 2017.

Qualitative Multitask Learning Results



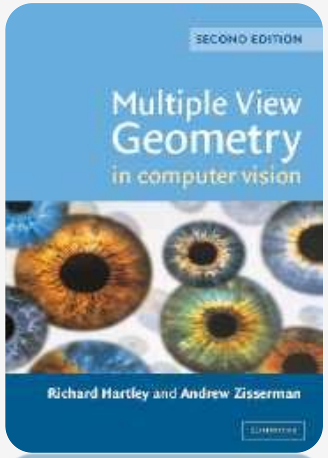
(a) Input image (b) Segmentation output (c) Instance output (d) Depth output

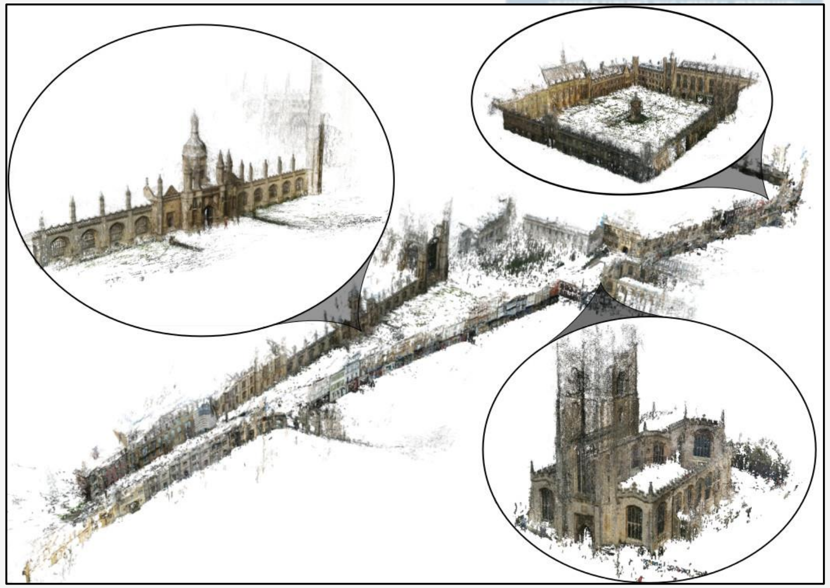
[1] Alex Kendall, Yarin Gal and Roberto Cipolla. **Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics.** arxiv preprint 1705.07115, 2017.

Geometry

Geometry in Computer Vision?

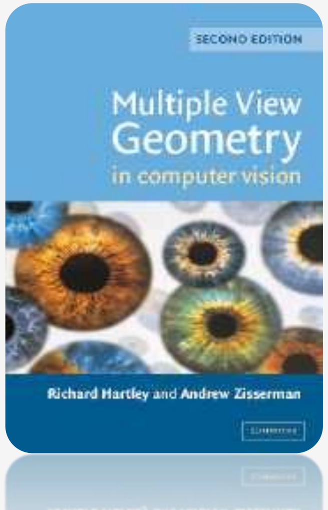
- Geometry was once the most exciting topic in computer vision
- Now machine learning models are the solution to most tasks
- These black boxes can learn many representations with end-to-end supervised learning
- Often naïve architectures are used

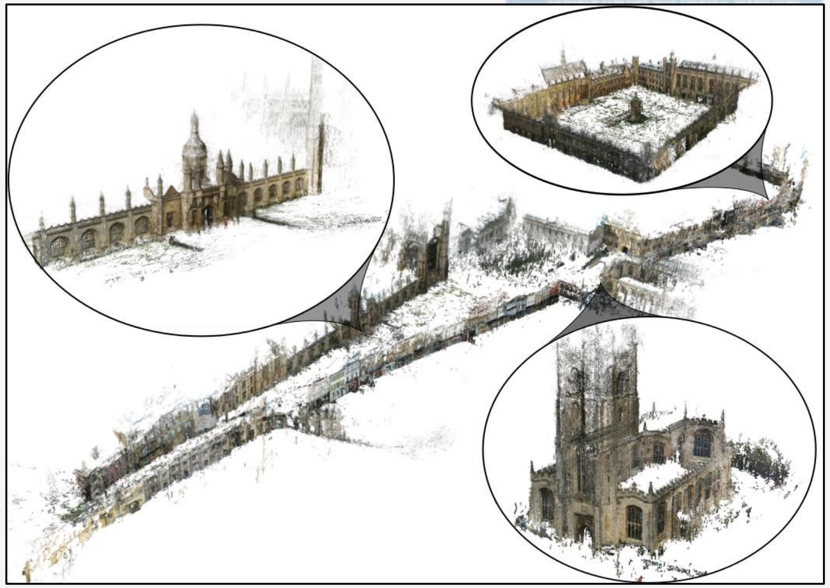




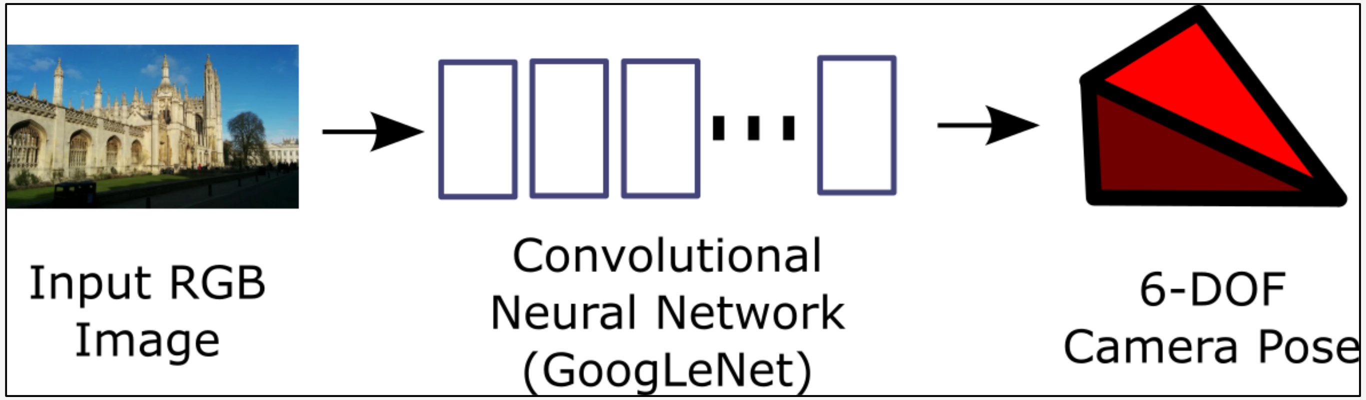
Geometry in Computer Vision?

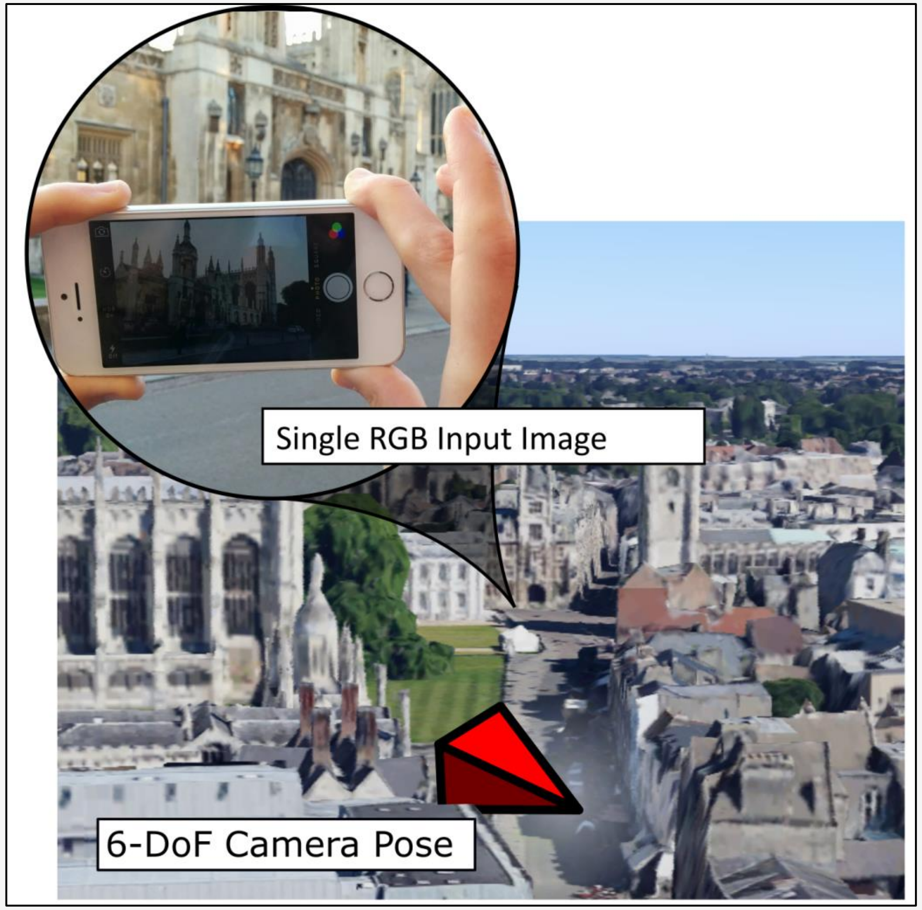
- *However*, geometry provides a rich source of training data
- Motion, pose and depth can be leveraged for supervised and unsupervised training
- Geometric priors and architectural designs can significantly improve model performance





Naïve deep learning approach to learning camera pose





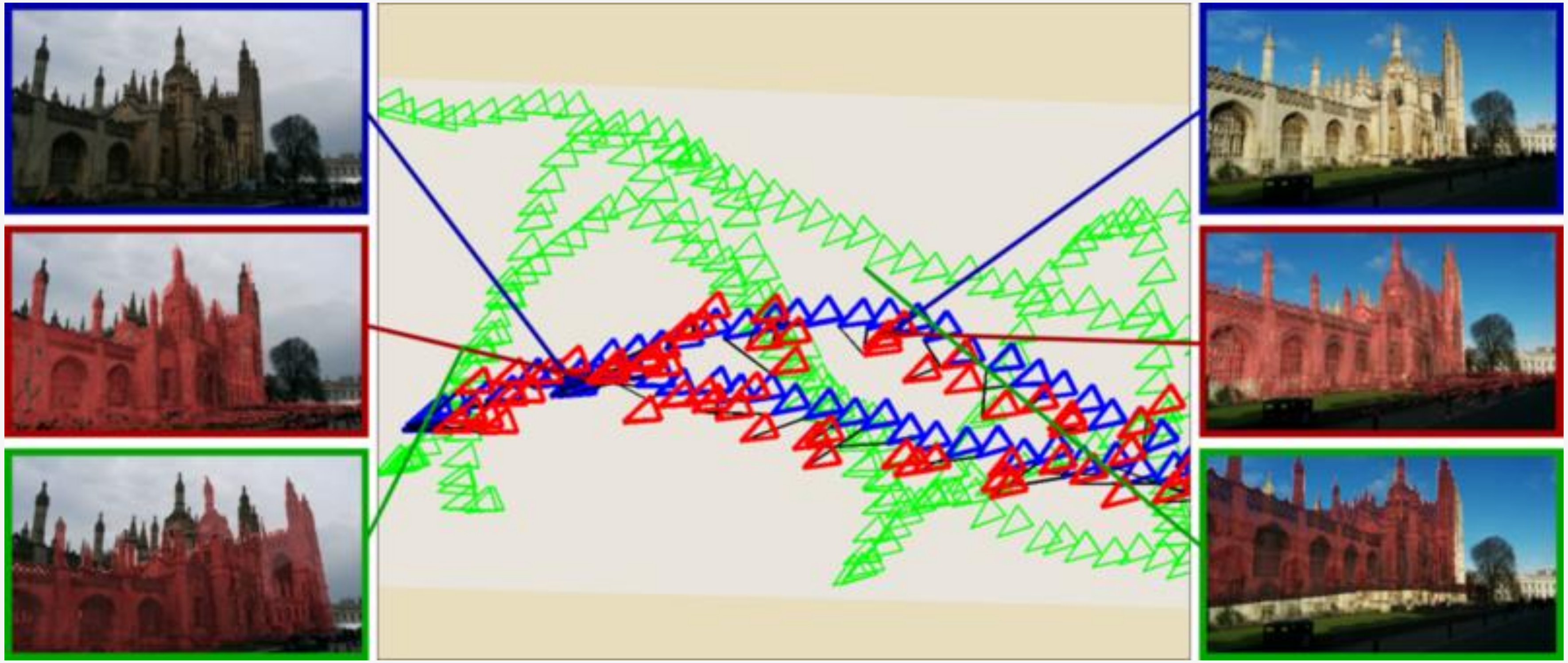
PoseNet: trained end-to-end to regress camera position, *x* and orientation, *q*

$$loss(I) = \|\hat{x} - x\|_2 + \beta \left\| \hat{q} - \frac{q}{\|q\|} \right\|_2$$

Alex Kendall, Matthew Grimes and Roberto Cipolla. **PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization.** ICCV, 2015.

Camera Pose Regression

training data in green, test data in blue, PoseNet results in red



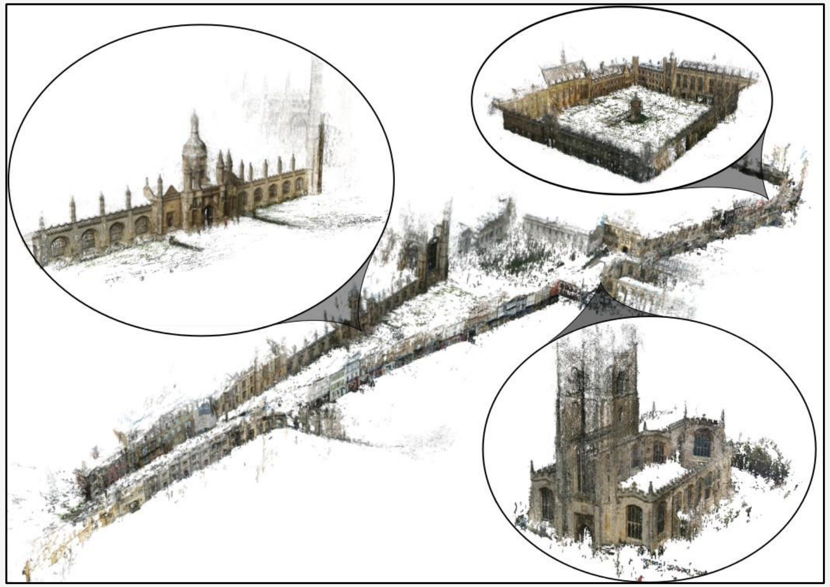
Alex Kendall, Matthew Grimes and Roberto Cipolla. **PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization.** ICCV, 2015.

Learning camera pose, *with geometry*

Train with reprojection loss of 3-D geometry

with predicted and ground truth camera poses.

$$loss(I) = \frac{1}{|\mathcal{G}'|} \sum_{g_i \in \mathcal{G}'} \left\| \pi(\mathbf{q}, \mathbf{x}, \mathbf{g_i}) - \pi(\hat{\mathbf{q}}, \hat{\mathbf{x}}, \mathbf{g_i}) \right\|_{\gamma}$$



Where π is the projection function of 3-D point g_i

Alex Kendall and Roberto Cipolla. **Geometric loss functions for camera pose regression with deep learning.** CVPR, 2017 (*to appear*).

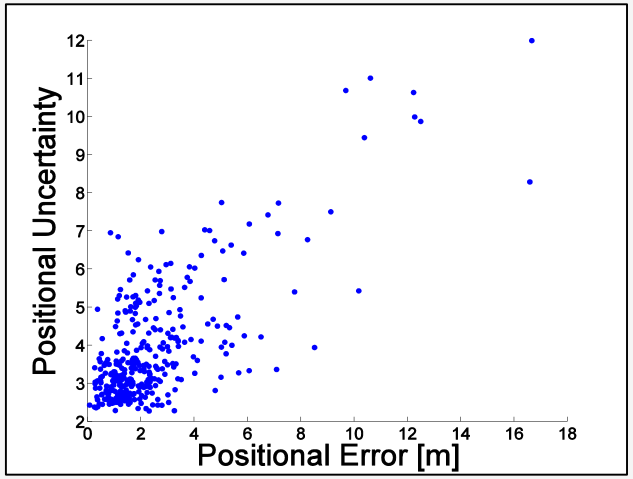
Camera Pose Regression

Using geometry in our model structure improves performance

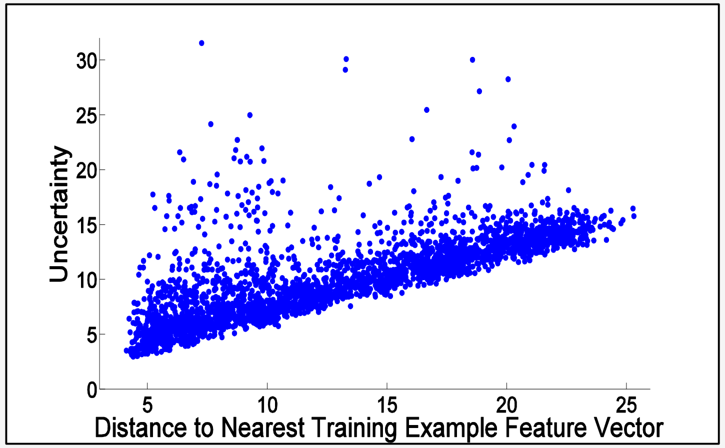
Scene	Spatial Extent	PoseNet (GoogLeNet, L2) [20]	Bayesian PoseNet (GoogLeNet, L2) [19]	PoseNet v2 (this work) (ResNet, L1+reprojection)
King's College	140 × 40m	1.66m, 4.86°	1.74m, 4.06°	0.92m, 0.83°
Street	500 × 100m	2.96m, 6.00°	2.14m, 4.96°	1.32m, 1.57°
Old Hospital	50 × 40m	2.62m, 4.90°	2.57m, 5.14°	1.12m, 1.83°
Shop Façade	35 × 25m	1.41m, 7.18°	1.25m, 7.54°	0.72m, 0.93°
St Mary's Church	80 × 60m	2.45m, 7.96°	2.11m, 8.38°	1.62m, 1.84°
Average		2.22m, 6.18°	1.96m, 6.02°	1.14m, 1.40°
Chess	3×2×1m	0.32m, 6.60°	0.37m, 7.24°	0.12m, 3.24°
Fire	2.5×1× 1m	0.47m, 14.0°	0.43m, 13.7°	0.13m, 4.20°
Heads	2×0.5×1m	0.30m, 12.2°	0.31m, 12.0°	0.08m, 5.72°
Office	2.5×2×1.5m	0.48m, 7.24°	0.48m, 8.04°	0.16m, 2.38°
Pumpkin	2.5×2×1m	0.49m, 8.12°	0.61m, 7.08°	0.14m, 2.15°
Red Kitchen	4×3×1.5m	0.58m, 8.34°	0.58m, 7.54°	0.16m, 4.24°
Stairs	2.5×2×1.5m	0.48m, 13.1°	0.48m, 13.1°	0.18m, 4.86°
Average		0.45m, 9.94°	0.47m, 9.81°	0.14m, 3.83°

Alex Kendall and Roberto Cipolla. **Geometric loss functions for camera pose regression with deep learning.** CVPR, 2017 (*to appear*).

Epistemic uncertainty to estimate loop closure



We can use epistemic uncertainty to estimate metric relocalisation error



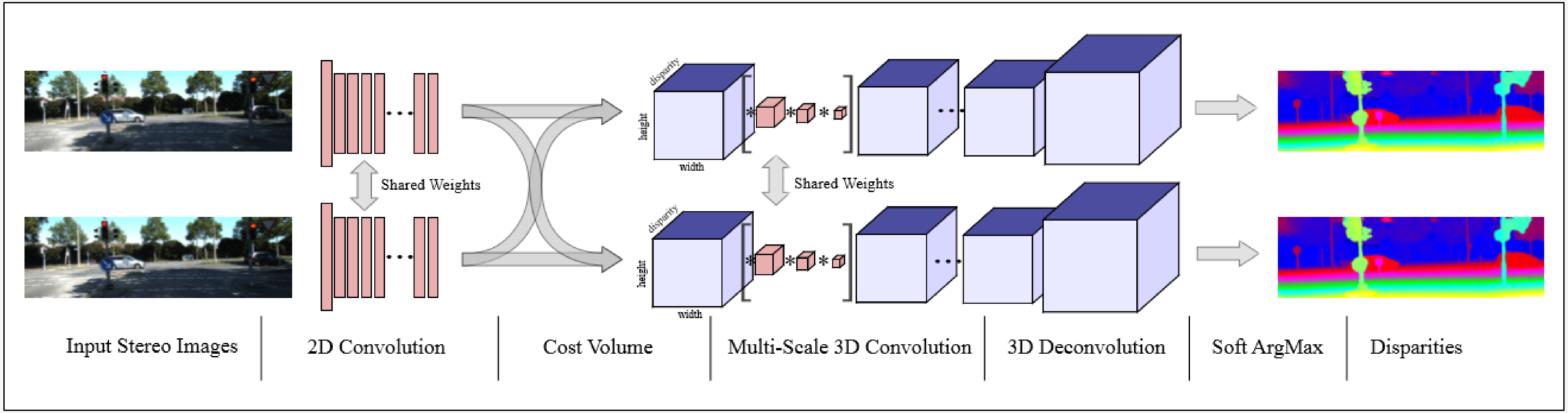
Determine if the model has seen the landmark before (loop closure)

Increased uncertainty from strong occlusion, motion blur, visually ambiguous landmarks

Alex Kendall and Roberto Cipolla. **Modelling Uncertainty in Deep Learning for Camera Relocalization.** ICRA, 2016.

End to end deep learning for stereo vision

- Form differentiable cost volume and sub-pixel regression network with soft argmax function
- Use 3-D convolutions to learn to regularise the volume

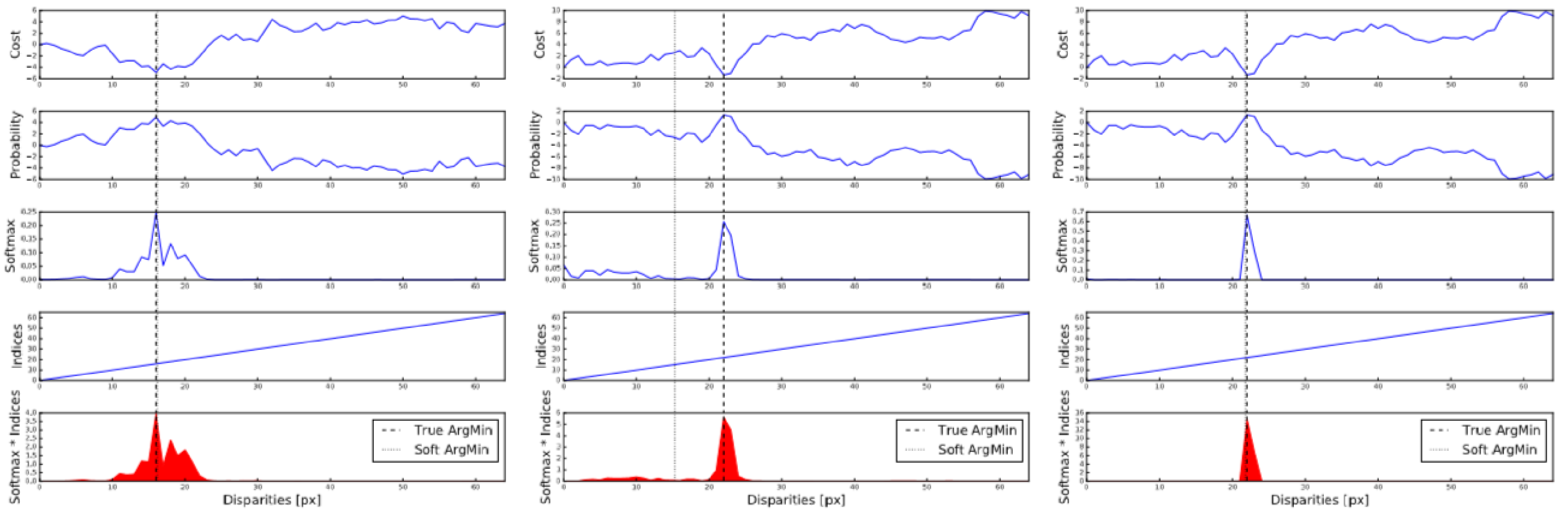


Alex Kendall et al. **End-to-End Learning of Geometry and Context for Deep Stereo Regression.** arXiv preprint 1703.04309, 2017.

Soft ArgMin / ArgMax

$$soft\,argmin := \sum_{d=0}^{D_{max}} d \times \sigma(-c_d)$$

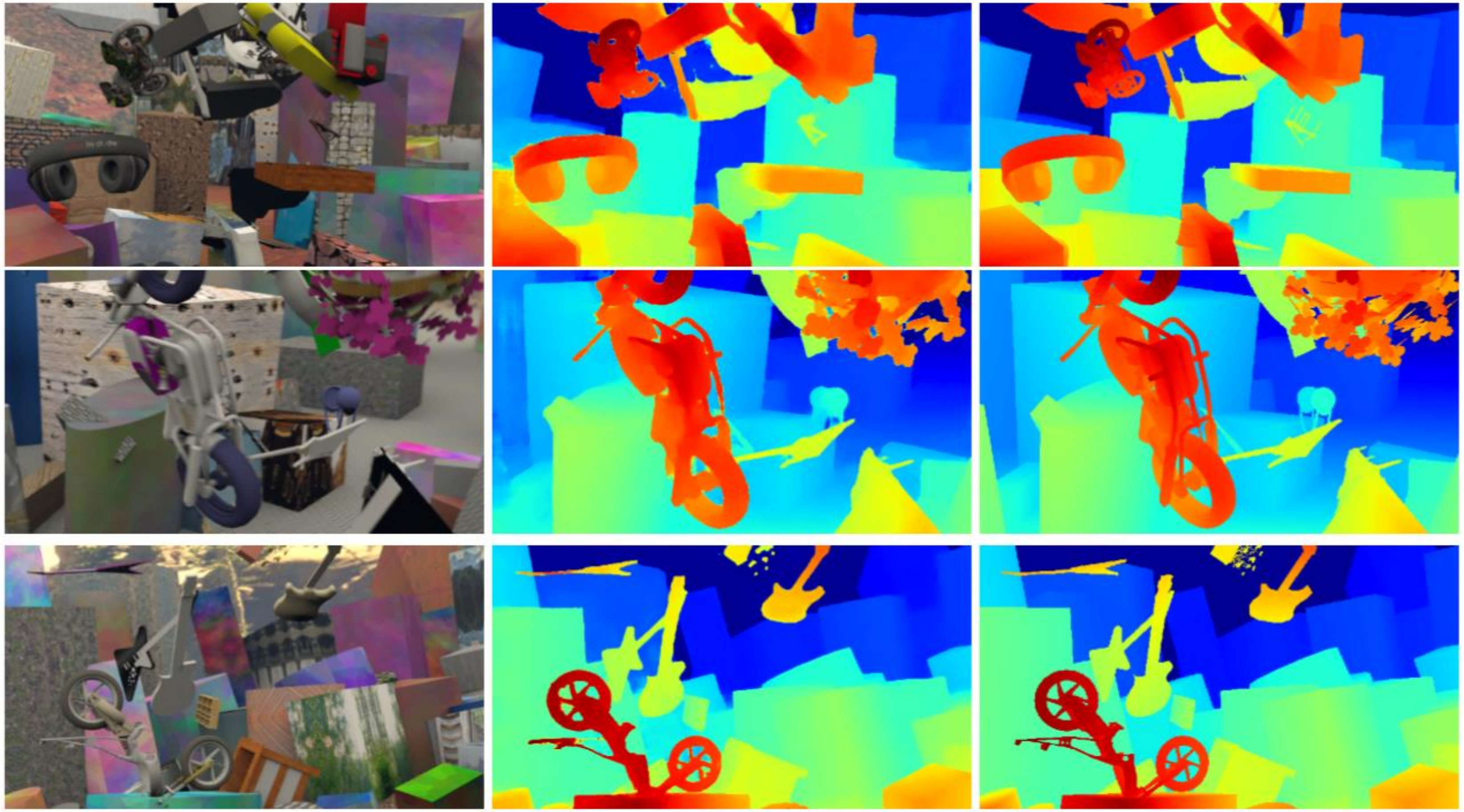
Loss	> 3px Error
Classification Loss	12.2
Soft Classification	12.3
Regression	9.34



(a) Soft ArgMin (b) Multi-modal distribution (c) Multi-modal distribution with prescaling

Alex Kendall et al. **End-to-End Learning of Geometry and Context for Deep Stereo Regression.** arXiv preprint 1703.04309, 2017.

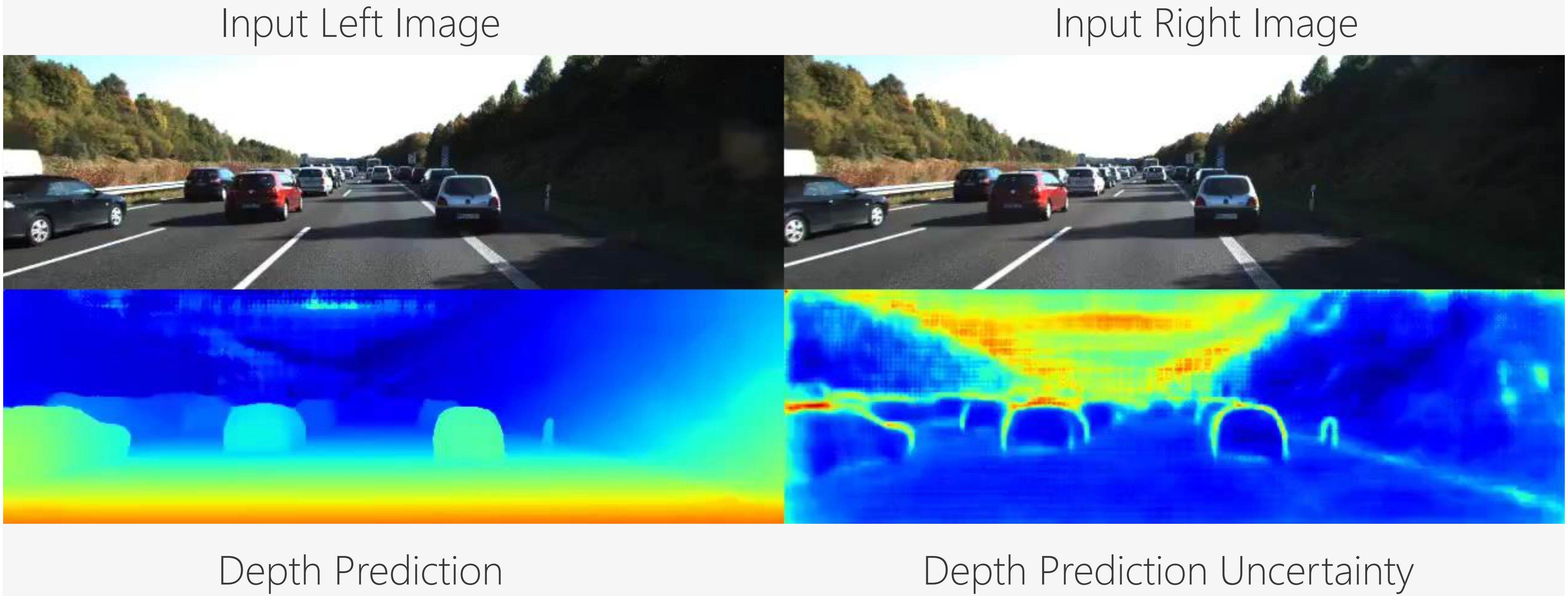
Scene Flow Dataset Results



(c) Scene Flow test set qualitative results. From left: left stereo input image, disparity prediction, ground truth.

Alex Kendall et al. **End-to-End Learning of Geometry and Context for Deep Stereo Regression.** arXiv preprint 1703.04309, 2017.

Probabilistic Deep Learning for Stereo Vision



Alex Kendall et al. **End-to-End Learning of Geometry and Context for Deep Stereo Regression.** arXiv preprint 1703.04309, 2017.

Alex Kendall and Roberto Cipolla. **Uncertainty and Unsupervised Learning for Stereo Vision with Probabilistic Deep Learning.** *Under Review*, 2017.

1st Place on the 2012 & 2015 KITTI Stereo Challenge

The KITTI Vision Benchm ×

www.cvlibs.net/datasets/kitti/eval_scene_flow.php?benchmark=stereo

The KITTI Vision Benchmark Suite

A project of Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago

Karlsruhe Institute of Technology

home setup stereo flow scene flow odometry object tracking road semantics raw data submit results jobs

Andreas Geiger (MPI Tübingen) | Philip Lenz (KIT) | Christoph Stiller (KIT) | Raquel Urtasun (University of Toronto)

Stereo Evaluation 2015

	Method	Setting	Code	D1-bg	D1-fg	D1-all	Density	Runtime	Environment	Compare
1	GC-NET			2.21 %	6.16 %	2.87 %	100.00 %	0.9 s	Nvidia GTX Titan X	
A. Kendall, H. Martirosyan, S. Dasgupta, P. Henry, R. Kennedy, A. Bachrach and A. Bry: End-to-End Learning of Geometry and Context for Deep Stereo Regression. arXiv preprint arXiv:1703.04309 2017.										
2	DRR			2.58 %	6.04 %	3.16 %	100.00 %	0.4 s	Nvidia GTX Titan X	
3	L-ResMatch		code	2.72 %	6.95 %	3.42 %	100.00 %	48 s	1 core @ 2.5 Ghz (C/C++)	
A. Shaked and L. Wolf: Improved Stereo Matching with Constant Highway Networks and Reflective Loss. arXiv preprint arxiv:1701.00165 2016.										
4	Displets v2		code	3.00 %	5.56 %	3.43 %	100.00 %	265 s	>8 cores @ 3.0 Ghz (Matlab + C/C++)	
F. Guney and A. Geiger: Displets: Resolving Stereo Ambiguities using Object Knowledge. Conference on Computer Vision and Pattern Recognition (CVPR) 2015.										
5	CNNF+SGM			2.78 %	7.69 %	3.60 %	100.00 %	71 s	TESLA K40C	
6	PBCP			2.58 %	8.74 %	3.61 %	100.00 %	68 s	Nvidia GTX Titan X	
A. Seki and M. Pollefeys: Patch Based Confidence Prediction for Dense Disparity Map. British Machine Vision Conference (BMVC) 2016.										
7	SN			2.66 %	8.64 %	3.66 %	100.00 %	67 s	Titan X	

Alex Kendall et al. **End-to-End Learning of Geometry and Context for Deep Stereo Regression.** arXiv preprint 1703.04309, 2017.

SKYDIO

Autonomous Drone Prototype

Conclusions

1 *Aleatoric* uncertainty is important for

- **Large data situations**, where epistemic uncertainty is explained away,
- **Real-time applications**, because we can form aleatoric models without expensive Monte Carlo samples,
- **Multitask applications**, because we can appropriately weight each loss.

2 *Epistemic* uncertainty is important for

- **Safety-critical applications**, because epistemic uncertainty is required to understand examples which are different from training data,
- **Small datasets**, where the training data is sparse,
- **Exploratory applications**, such as loop closure and reinforcement learning.

Conclusions

3. *It is important to quantify the accuracy of uncertainty estimates*

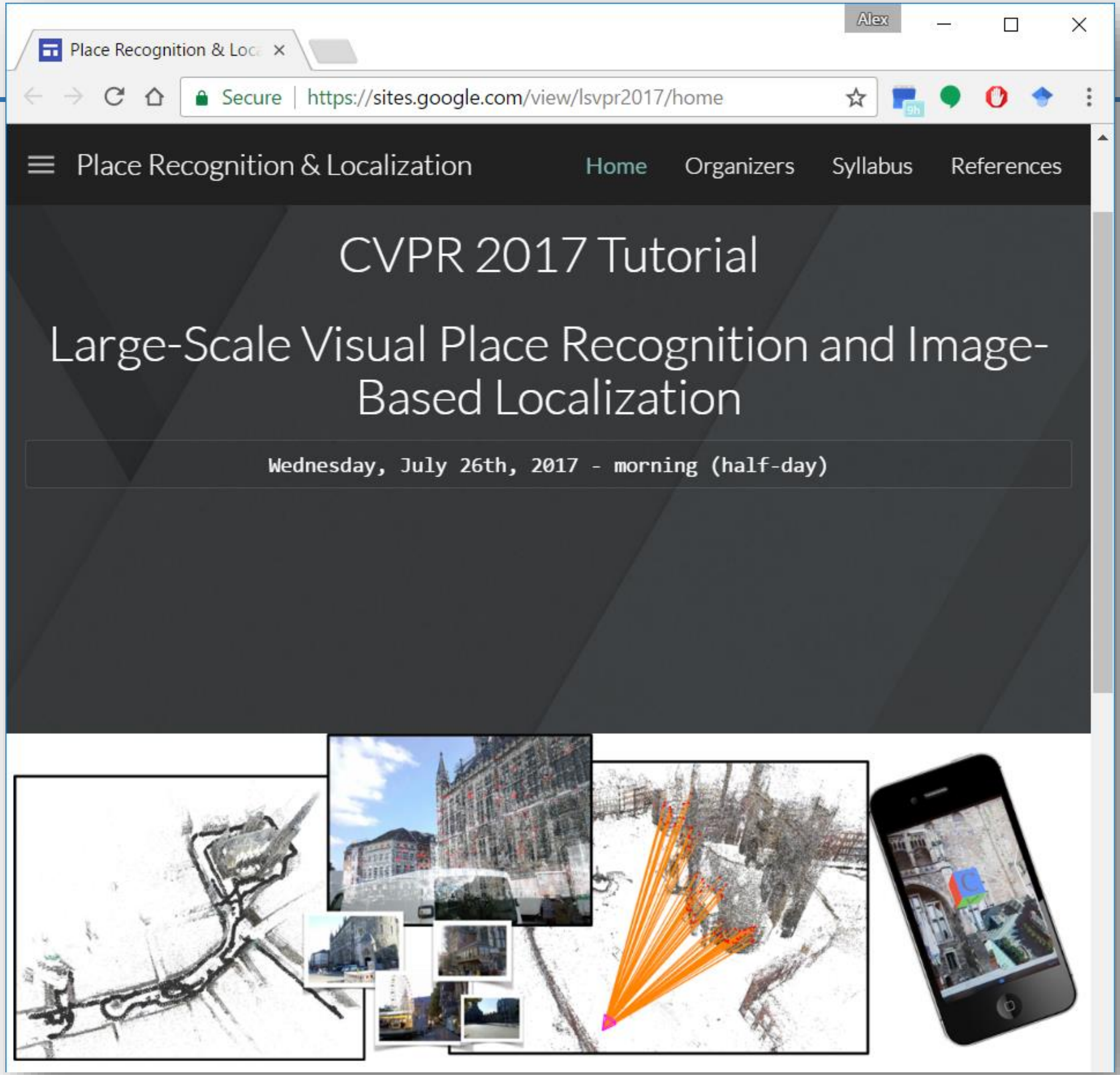
4. *We should leverage our knowledge of geometry when designing machine learning models for computer vision*
   - *Reprojection loss*
   - *Stereo cost volume*

CVPR Tutorial

Hawaii

July 26th 2017

See you there?



Thank You & References

- Alex Kendall and Yarin Gal. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? arXiv preprint 1703.04977, 2017.
- Alex Kendall and Roberto Cipolla. Geometric loss functions for camera pose regression with deep learning. CVPR, 2017 (*to appear*).
- Alex Kendall et al. End-to-End Learning of Geometry and Context for Deep Stereo Regression. arXiv preprint 1703.04309, 2017.
- Alex Kendall and Roberto Cipolla. Uncertainty and Unsupervised Learning for Stereo Vision with Probabilistic Deep Learning. *Under Review*, 2017.
- Alex Kendall et al. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. arxiv preprint 1705.07115, 2017.
- Vijay Badrinarayanan, Alex Kendall and Roberto Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. PAMI 2017.
- Alex Kendall and Roberto Cipolla. Modelling Uncertainty in Deep Learning for Camera Relocalization. ICRA, 2016.
- Alex Kendall et al. Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding. arXiv 1511.02680, 2015.
- Alex Kendall, Matthew Grimes and Roberto Cipolla. PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization. ICCV, 2015.