Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

R&D Project Proposal

Uncertainty estimation for Key-point detection task : Robustness study

Nandhini Shree Mathivanan

Supervised by

Prof.Nico Hochgeschwender
Mr.Deebul Nair0098765

April 2022

1 Introduction

Deep learning is an artificial intelligence (AI) function that mimics the human brain's processing and pattern-making processes to aid decision-making. Key point detection is one such pattern-making process used by the brain to make decisions. A keypoint or feature in an image which can be defined as an unique meaningful structure in that image, however it is semantically ill-defined, in the sense that it is unclear what keypoints are relevant for any given input image.More advanced computer vision tasks such as structure from motion, object recognition, 3D reconstruction, simultaneous localization and mapping (SLAM), content-based retrieval and image matching depends on keypoint detection and description methods[7].
New keypoint detection and description approaches have arisen as a result of the advent of deep learning methods, particularly convolutional models,as they are claiming to outperform traditional algorithms on benchmarks. Deep learning models are expected to learn abstract image features from large datasets. Convolutional models learn about features by supervision rather than handcrafting them, hence their performance is strongly reliant on ground truth data.
In everyday scenarios, we deal with uncertainties in numerous fields, from invest opportunities and medical diagnosis to sporting games and weather forecast, with an objective to make decision based on collected observations and uncertain domain knowledge[14]. Understanding the uncertainty of a neural network's (NN) predictions is essential for many applications. Two types of uncertainty are often of interest: epistemic uncertainty, which is inherent to the model, caused by a lack of training data and aleatoric uncertainty which is caused by inherent noise and ambiguity in data and hence irreducible[9].

An unbiased evalution over state-of-the-art keypoint detection and the proposed methods using key-point detection datasets, like face recognition or object detection, is crucial. Hence in this research we focus on robustness study for uncertainty estimation for key-point detection task.

Why is it important?

It is important to identify uncertainties as it plays a crucial role in risk assessment and decision making process. When an inaccurate measurement results are detected it leads to increase in decision risks. In case of safety-critical applications relying on Neural Network models, a measure of uncertainty can be helpful to decide whether low-confidence decisions need further validation. The self-driving software is a key differentiating feature of highly automated vehicles. The software is based on machine learning algorithms and deep learning neural networks that include millions of virtual neurons that mimic the human brain.
In May 2016 there was the first fatality from an assisted driving system, caused by the perception system confusing the white side of a trailer for bright sky [33]. In a second recent example, an image classification system erroneously identified two African Americans as gorillas [11], raising concerns of racial discrimination. If both these algorithms were able to assign a high level of uncertainty to their erroneous predictions, then the system may have been able to make better decisions and likely avoid disaster.
In this project we focus on a high-dimensional regression task such as keypoint detection task. Keypoint detection entails detecting people while also locating their keypoints. Interest points and keypoints are the same thing. They are spatial positions in the image or points in the image that determine what is fascinating or stands out in an image. They are invariant to image rotation,translation, shrinkage, distortion, and so on.
Uncertainty estimation methods will give an overview and help other researchers to choose the best suitable uncertainty estimation methods for both regression and classification models depending on their requirements. At the same time, It also helps to promotes research related to this topic.

1.1 Problem statement

The output of DNNs are stochastic in nature which involves randomness and uncertainties. In high-risk applications, identifying such uncertainties plays a very crucial role. Deep neural networks (DNN) have become popular in safety-critical and autonomous systems because of their capacity to estimate uncertainty along

with network prediction. Accurate and calibrated uncertainty can be used to build confidence in autonomous systems decision-making. Because there is no ground truth for uncertainty, uncertainty estimation is a challenging task, especially in high-dimensional data. The existence of noisy labels or outliers in the training data makes uncertainty estimation more challenging.

1.2 Research questions

There are several methods which can do uncertainty estimation, In this research we improve the robustness of uncertainty estimation by modeling the loss function using a heavy-tailed distribution and robustness of the loss function is evaluated using the complex regression tasks such as Keypoint estimation.
Furthermore, this work aims to answer the following research questions.

- **Survey on uncertainty estimation methods for Keypoint detection?**
  - What are the different papers in keypoint detection?
  - Which are the different tasks solved using keypoint detection?
  - What datasets are available for each keypoint detection task?
  - What loss functions are used for keypoint detection task ?
- **Comparison of uncertainty estimation methods for Keypoint detection?**
- **Robustness study on uncertainty estimation methods for Keypoint detection**

2 Related Work

Uncertainty estimation for regression:
In deep learning one of the most preferred approaches for uncertainty estimation is Bayesian approach and ensembles. In Bayesian approach the weights of the neural networks are replaced by the parametric distributions and it had

prohibitively high computational cost. In deep ensembles instead of getting single output we get a set of distributional parameters.
Uncertainty Estimation methods also differ based on number of forward passes required by an approach. Methods such as [22] require only a single forward pass whereas methods like [2] require several passes of an input through a network to estimate uncertainty.
Most approaches for estimating uncertainty in deep learning based on ensembling or Monte Carlo sampling. A new method is introduced to estimate uncertainty in a single forward pass. [35] As a result, even after human annotation, real-world datasets contain significant label noise

Robustness study Uncertainty estimation for regression:

For regression tasks, this paper [27] proposes the use of a heavy-tailed distribution (Laplace distribution) to improve the robustness to outliers. This property is evaluated using standard regression benchmarks and on a high-dimensional regression task of monocular depth estimation, both containing outliers. This paper [2] demonstrate learning well-calibrated measures of uncertainty on various benchmarks, scaling to complex computer vision tasks, as well as robustness to adversarial and OOD test samples.A new model [**?**] for estimating uncertainty in a single forward pass that works on both classification and regression problems. Combining bi-Lipschitz feature extractor with point approximate Gaussian process this results in good robustness and principled uncertainty estimation. Proposed method can allow changes in Deep kernel Learning(DKL) to match with softmax neural networks accuracy. This method overcomes previous work's limitations by addressing determining uncertainty quantification. They demonstrate the performance of DUE on regression on personalized healthcare.

Keypoint detection datasets:

The process of locating key object elements is known as KeyPoint detection. The eyebrows, nose tips and eye corners, for example, are important features of our faces. These components aid in the representation of the underlying object in a feature-rich way. Pose estimation, face detection, and more applications of KeyPoint detection exists in the KeyPoint detection datasets.Some of the common keypoint datasets are explained in this section.

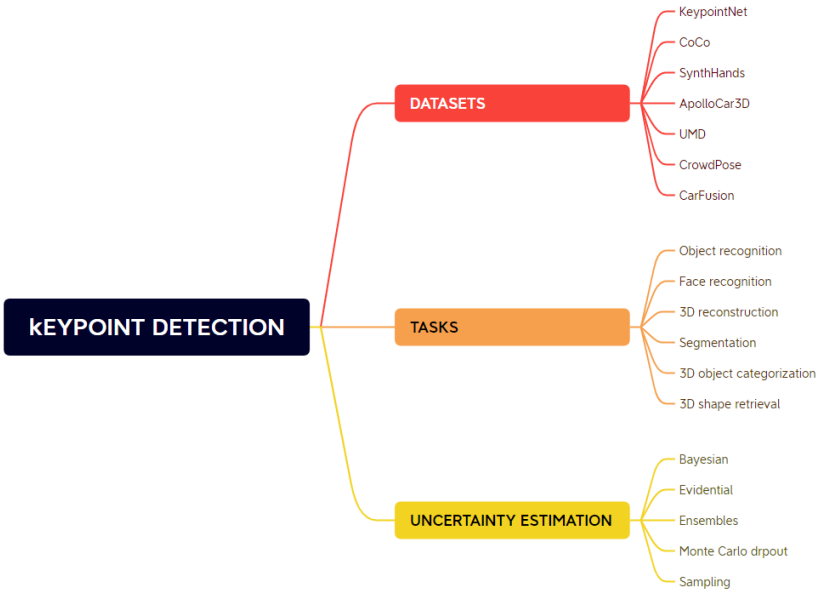

Figure 1: Mindmap for related work

By utilising multiple human annotations, KeypointNet is a large-scale and diverse 3D keypoint dataset that comprises 83,231 keypoints and 8,329 3D models from 16 item categories. It is based on ShapeNet models and contains 8,329 3D models,83,231 keypoints from 16 object categories.
The MS COCO dataset (Microsoft Common Objects in Context) is a large-scale dataset for segmentation,object detection, captioning and key-point detection. There are 328K photos in the dataset.detection of keypoints: about 250,000 human instances and200,000 photos have been tagged with keypoints (17 possible keypoints, such as nose,left eye,right ankle,right hip)
The SynthHands dataset contains genuine captured hand motion that has been retargeted to a virtual hand with natural backgrounds and interactions with various objects for hand posture estimation. The dataset includes data for both female and male hands, both with and without object engagement. The hand and foreground object were synthesized using Unity. Realistic item textures and background images (in terms of depth and color) were also utilised. For 21 keypoints on the hand, ground truth 3D locations are

presented.
The face dataset UMDFaces is divided into two parts:
367,888 facial annotations for 8,277 people in still images
Over 3.7 million video frames which are annotated from over 22,000 videos featuring 3100 subjects.

Keypoint detection tasks:

The computation of similarity between three dimensional (3D) surfaces is key to a number of pattern recognition tasks such as 3D modeling and 3D object recognition [23]. The aim of 3D modeling is to measure the similarity between 3D surfaces captured from different viewpoints, align and merge them to construct a complete 3D model of an object [8]. On the other hand, the task of 3D object recognition consists in correctly determining the identity and pose of objects in a scene[30]. Both these tasks find vast applications in fields such as robotics [29], reverse engineering [39], scene understanding [36], medical and biometric systems [18].
Human keypoint detection is also known as human pose estimation (HPE) refers to detecting human body keypoint location and recognizing their categories for each person instance from a given image. It is very useful in many downstream applications such as activity recognition[21], human–robot interaction [25], and video surveillance[12].

Keypoint detection architecture:

2D keypoint detection has become a popular research topic these years for its wide use in computer vision applications. This paper [34] propose a multi-resolution framework that generates heatmaps representing per-pixel likelihood for keypoints. Hourglass [14] develops a repeated bottom-up, top-down architecture, and enforces intermediate supervision by applying loss on intermediate heatmaps. On the other hand, the author [40] propose a simple and effective model that adds a few deconvolutional layers on ResNet. HRNet [31] maintains high resolution through the whole network and achieves notable improvement.

Uncertainty estimation for keypoint detection:

In this paper [20] they explore maximum likelihood estimation (MLE) to develop an efficient and effective regression-based methods. They propose

a novel regression paradigm with Residual Log-likelihood and benchmark the proposed method on three pose estimation datasets, including MPII , MSCOCO and Human3.6M. Estimation (RLE) to capture the underlying output distribution. In the paper [6]they tackle the fundamentally ill-posed problem of 3D human localization from monocular RGB images. Driven by the limitation of neural networks outputting point estimates,they address the ambiguity in the task by predicting confidence intervals through a loss function based on the Laplace distribution. The architecture is a light-weight feed-forward neural network that predicts 3D locations and corresponding confidence intervals given 2D human poses.

Heavy tailed distribution:

A heavy tailed distribution has a tail that's heavier than an exponential distribution. In other words, a distribution that is heavy tailed goes to zero slower than one with exponential tails; there will be more bulk under the curve of the PDF. Heavy tailed distributions tend to have many outliers with very high values. The heavier the tail, the larger the probability that you'll get one or more disproportionate values in a sample[15]

Metrics:

The most commonly used metrics are Root Mean Squared Error(RMSE), Negative-Log Likelihood(NLL), Explained variance, etc. In this paper [4] the methods are evaluated with Root Mean Squared Error (RMSE), Log Likelihood (LL) and Tail Calibration Error (TCE). The proposed evidential regression method [2] against results presented in this paper for model ensembles and dropout is based on root mean squared error (RMSE), negative log-likelihood (NLL), and inference speed.

3 Project Plan

3.1 Work Packages

The bare minimum will include the following packages:

WP1 Literature review Uncertainty estimation

- T1.1 Understanding uncertainty in Deep Learning(DL)
- T1.2 Literature search of uncertainty estimation methods in DL
- T1.3 Survey on Keypoint detection datasets
- T1.4 Identify and analyse the state-of-the-art methods

WP2 Experimental setup and analysis

- T2.1 Choosing the best uncertainty estimation methods
- T2.2 Implementation of the selected method with collected datasets
- T2.3 Validation of implemented methods based on robustness.
- T2.4 Validation of implemented methods

WP3 Mid-term Report

- T3.1 Submission of the Mid-term report

WP4 Benchmarking

- T4.1 Benchmark robustness of uncertainty estimation methods

WP5 Project Report

- T5.1 Revision
- T5.2 Final submission of the report

3.2 Milestones

M1 Literature search

M2 Experimental setup and analysis

M3 Benchmarking

M4 Report submission

3.3 Project Schedule

3.4 Deliverables

Minimum Viable

- Regression:
  - State of art analysis
  - Survey of data sets
  - Survey on metrics
  - Comparison of 2 uncertainty estimation methods on a keypoint detection dataset

Expected

- Regression:
  - Comparison of multiple uncertainty estimation methods on multiple keypoint detection datasets
  - Analysing the robustness of the state-of-the-art methods .

Desired

- Regression:
  - Benchmark robustness of uncertainty estimation methods using heavy-tail distribution

References

[1] MS Windows NT kernel description. `https://www.edureka.co/blog/classification-in-machine-learning/`.

[2] Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. Deep evidential regression. *NeurIPS*, 2019.

[3] Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. Deep evidential regression. *Advances in Neural Information Processing Systems*, 33: 14927–14937, 2020.

[4] Javier Antorán, James Allingham, and José Miguel Hernández-Lobato. Depth uncertainty in neural networks. *Advances in neural information processing systems*, 33:10620–10634, 2020.

[5] Eric Arazo, Diego Ortego, Paul Albert, Noel O'Connor, and Kevin McGuinness. Unsupervised label noise modeling and loss correction. In *International conference on machine learning*, pages 312–321. PMLR, 2019.

[6] Lorenzo Bertoni, Sven Kreiss, and Alexandre Alahi. Monoloco: Monocular 3d pedestrian localization and uncertainty estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6861–6871, 2019.

[7] David Bojanić, Kristijan Bartol, Tomislav Pribanić, Tomislav Petković, Yago Diez Donoso, and Joaquim Salvi Mas. On the comparison of classic and deep keypoint detector and descriptor methods. In *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 64–69. IEEE, 2019.

[8] Do Hyun Chung, Il Dong Yun, and Sang Uk Lee. Registration of multiple-range views using the reverse-calibration technique. *Pattern Recognition*, 31 (4):457–464, 1998.

[9] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? does it matter? *Structural safety*, 31(2):105–112, 2009.

[10] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

[11] Jessica Guynn. Google photos labeled black people 'gorillas'. *USA Today*, 2015.

[12] Hironori Hattori, Namhoon Lee, Vishnu Naresh Boddeti, Fares Beainy, Kris M Kitani, and Takeo Kanade. Synthesizing a scene-specific pedestrian detector and pose estimator for static video surveillance. *International Journal of Computer Vision*, 126(9):1027–1044, 2018.

[13] Marton Havasi, Rodolphe Jenatton, Stanislav Fort, Jeremiah Zhe Liu, Jasper Snoek, Balaji Lakshminarayanan, Andrew M Dai, and Dustin Tran. Training independent subnetworks for robust prediction. *arXiv preprint arXiv:2010.06610*, 2020.

[14] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[15] Heavy-tail. Heavy tailed distribution light tailed distribution. URL `https://www.statisticshowto.com/heavy-tailed-distribution/`.

[16] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017.

[17] keypoint. Keypoint detection with transfer learning. URL `https://keras.io/examples/vision/keypoint_detection/`.

[18] Salman H Khan, M Ali Akbar, Farrukh Shahzad, Mudassar Farooq, and Zeashan Khan. Secure biometric template generation for multi-factor authentication. *Pattern Recognition*, 48(2):458–472, 2015.

[19] Stefan Lee, Senthil Purushwalkam, Michael Cogswell, David Crandall, and Dhruv Batra. Why m heads are better than one: Training a diverse ensemble of deep networks. *arXiv preprint arXiv:1511.06314*, 2015.

[20] Jiefeng Li, Siyuan Bian, Ailing Zeng, Can Wang, Bo Pang, Wentao Liu, and Cewu Lu. Human pose regression with residual log-likelihood estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11025–11034, 2021.

[21] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. Deep learning for generic object detection: A survey. *International journal of computer vision*, 128(2):261–318, 2020.

[22] Antonio Loquercio, Mattia Segu, and Davide Scaramuzza. A general framework for uncertainty estimation in deep learning. *IEEE Robotics and Automation Letters*, 5(2):3153–3160, 2020.

[23] Athanasios Mademlis, Petros Daras, Dimitrios Tzovaras, and Michael G Strintzis. 3d object retrieval using the 3d shape impact descriptor. *Pattern Recognition*, 42(11):2447–2459, 2009.

[24] Andrey Malinin. *Uncertainty estimation in deep learning with application to spoken language assessment*. PhD thesis, University of Cambridge, 2019.

[25] Osama Mazhar, Sofiane Ramdani, Benjamin Navarro, Robin Passama, and Andrea Cherubini. Towards real-time physical human-robot interaction using skeleton information and hand gestures. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1–6. IEEE, 2018.

[26] Jishnu Mukhoti, Andreas Kirsch, Joost van Amersfoort, Philip HS Torr, and Yarin Gal. Deterministic neural networks with appropriate inductive biases capture epistemic and aleatoric uncertainty. *arXiv preprint arXiv:2102.11582*, 2021.

[27] Deebul S Nair, Nico Hochgeschwender, and Miguel A Olivares-Mendez. Maximum likelihood uncertainty estimation: Robustness to outliers. *arXiv preprint arXiv:2202.03870*, 2022.

[28] Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, David Sculley, Sebastian Nowozin, Joshua V Dillon, Balaji Lakshminarayanan, and Jasper Snoek. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. *arXiv preprint arXiv:1906.02530*, 2019.

[29] Syed Afaq Ali Shah, Mohammed Bennamoun, and Farid Boussaid. Iterative deep learning for image set based face and object recognition. *Neurocomputing*, 174:866–874, 2016.

[30] Xiaohu Song, Damien Muselet, and Alain Trémeau. Affine transforms between image space and color space for invariant local descriptors. *Pattern recognition*, 46(8):2376–2389, 2013.

[31] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5693–5703, 2019.

[32] Sutskever I Bruna J Erhan D Goodfellow I Fergus R Szegedy C, Zaremba W. Intriguing properties of neural networks. *In: International Conference on Learning Representations (ICLR)*, 2014.

[33] National Highway Traffic Safety Administration Technical report, U.S. Department of Transportation. Tesla crash preliminary evaluation report. *NHTSA. PE 16-007*, 2017.

[34] Jonathan J Tompson, Arjun Jain, Yann LeCun, and Christoph Bregler. Joint training of a convolutional network and a graphical model for human pose estimation. *Advances in neural information processing systems*, 27, 2014.

[35] Joost Van Amersfoort, Lewis Smith, Yee Whye Teh, and Yarin Gal. Uncertainty estimation using a single deep deterministic neural network. In *International Conference on Machine Learning*, pages 9690–9700. PMLR, 2020.

[36] Kai Wang, Boris Babenko, and Serge Belongie. End-to-end scene text recognition. In *2011 International conference on computer vision*, pages 1457–1464. IEEE, 2011.

[37] Dave waters. https://blog.bitsathy.ac.in/predicting-the-future-isnt-magic-its-artificial-intelligence-part-1/. 2019.

[38] Yeming Wen, Dustin Tran, and Jimmy Ba. Batchensemble: an alternative approach to efficient ensemble and lifelong learning. *International Conference on Learning Representation*, 2020.

[39] John Williams and Mohammed Bennamoun. A multiple view 3d registration algorithm with statistical error modeling. *IEICE TRANSACTIONS on Information and Systems*, 83(8):1662–1670, 2000.

[40] Bin Xiao, Haiping Wu, and Yichen Wei. Simple baselines for human pose estimation and tracking. In *Proceedings of the European conference on computer vision (ECCV)*, pages 466–481, 2018.