UGLLI Face Alignment: Estimating Uncertainty with Gaussian Log Likelihood Loss

MITSUBISHI ELECTRIC Changes for the Better | MITSUBISHI ELECTRIC RESEARCH LABORATORIES, INC

Abhinav Kumar*[1,2], Tim K. Marks*[1], Wenxuan Mou*[1,3], Chen Feng[4], Xiaoming Liu[5]

[1]Mitsubishi Electric Research Labs (MERL); [2]University of Utah; [3]University of Manchester; [4]New York University; [5]Michigan State University

*Denotes equal contributions

1. Overview

Why uncertainty?

- Most of the current state-of-the-art face alignment methods predict landmark locations but do not model the uncertainty associated with the prediction.
- Knowing how uncertain the predictions are can be critical for downstream tasks.

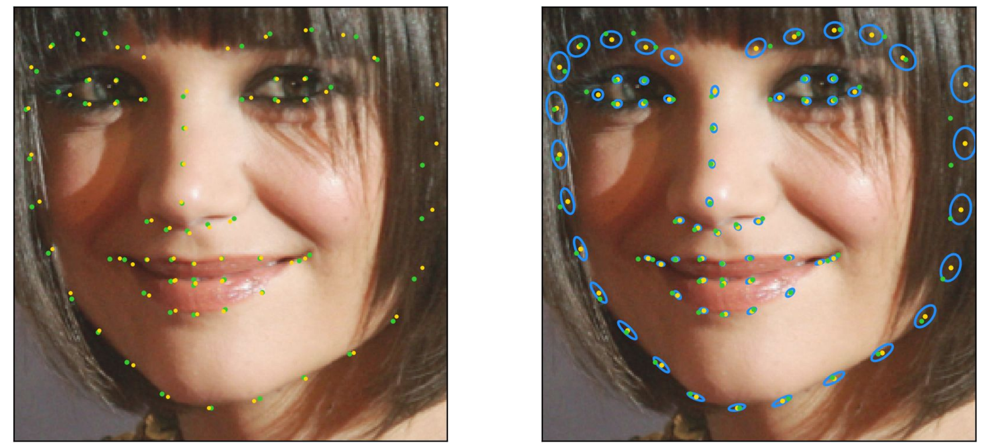


Figure: Difference between previous methods and our proposed method.

2. Issues with Previous Methods

1. Training: Use proxy ground truth heatmap

- Predict heatmap and use L_2 loss between predicted heatmap and a proxy ground truth heatmap:
  - Each proxy ground truth heatmap generated by placing a fixed symmetric Gaussian at the ground truth landmark location.
  - Heatmaps are not probabilities.

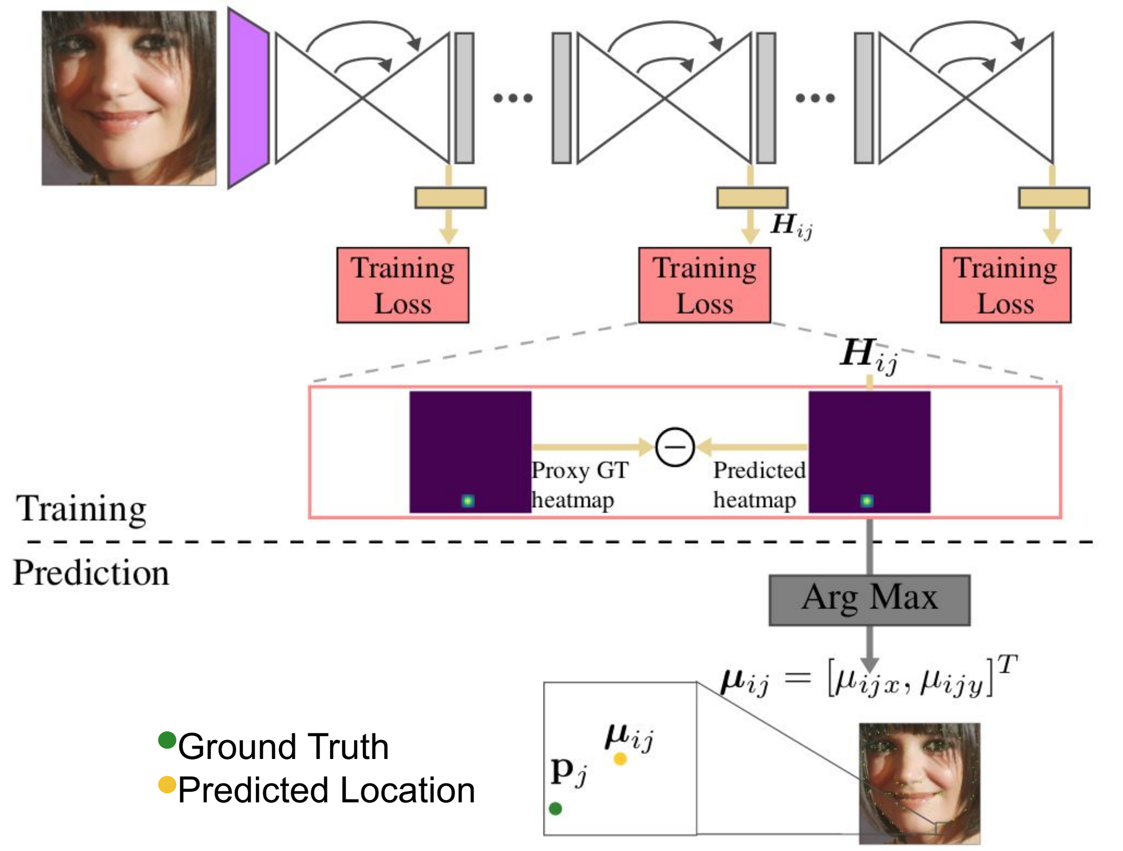


Figure: An example of stacked U-nets (stacked hourglass networks): DU-Net [2]. DU-Net also has dense connections among the hourglasses (not shown).

2. Prediction: Take arg max of the heatmap as the location estimate

- Accurate only up to one pixel. No sub-pixel accuracy.
- Training not end-to-end differentiable.

3. Why Not use Heatmaps to Model Probability Distributions?

- [1] uses a non-parametric mixture density network to learn the distribution as a mixture of a large number of Gaussians (a small symmetric Gaussian at each heatmap pixel).
- Our results show that the resolution of the heatmaps is not sufficient to accurately model uncertainty.

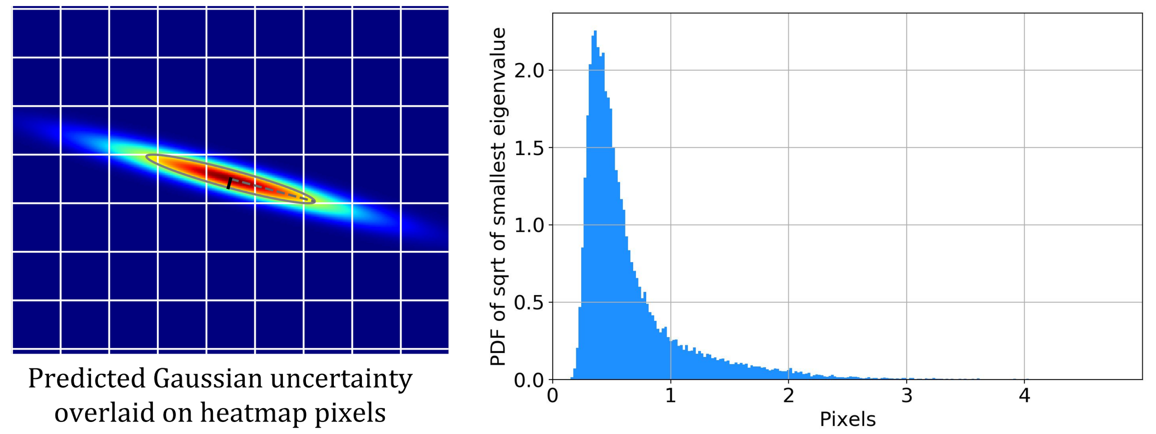


Figure:
- Histogram of square-root of smallest eigenvalue of Σ_{ij} (semi-minor axis of Gaussian ellipse [black line at left]).
- Gaussian is usually less than 1 heatmap pixel wide.
- Key reason heatmaps not suitable for accurate uncertainty estimation.

4. Proposed Method: UGLLI Face Alignment

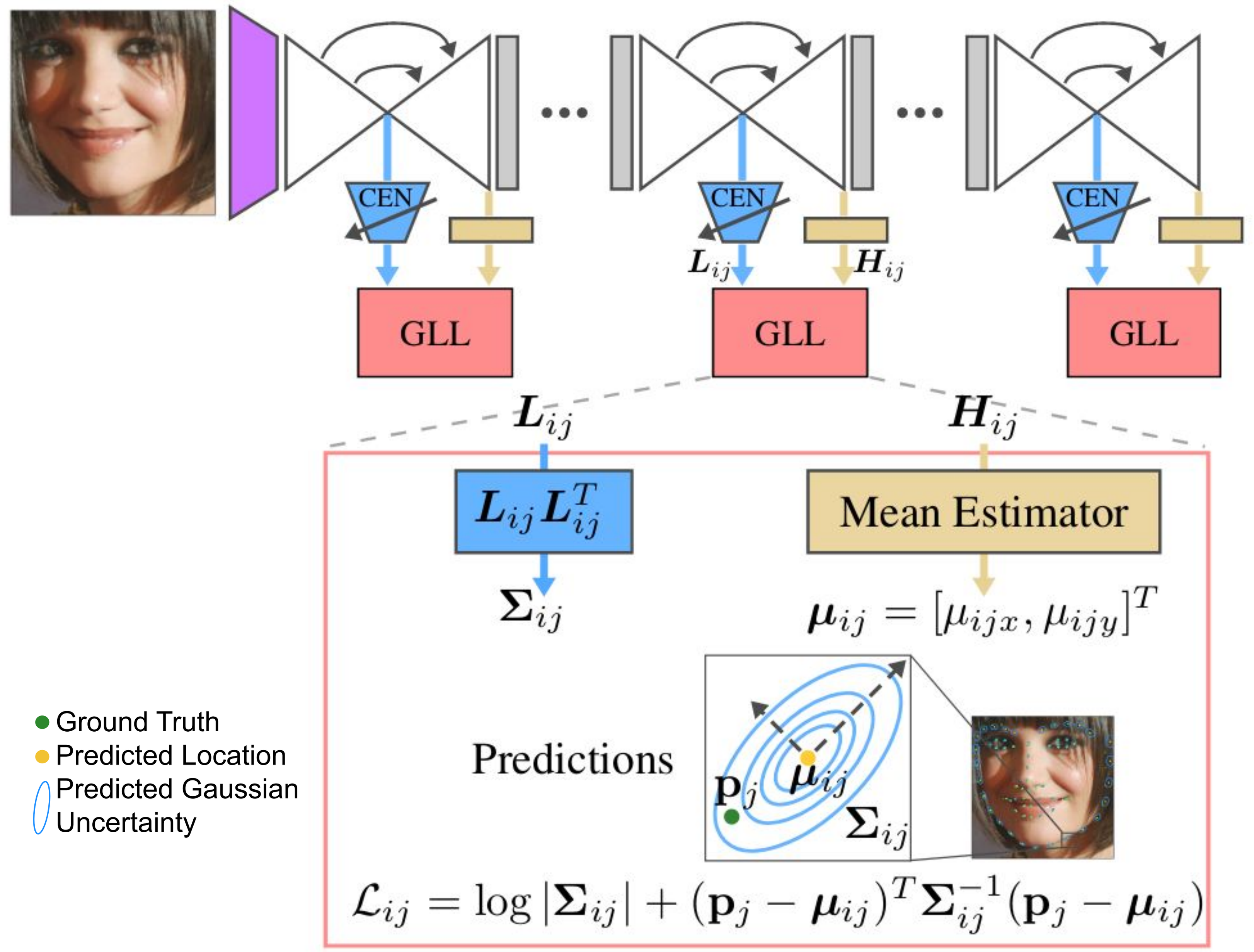


Figure: An overview of UGLLI Face Alignment with DU-Net [2] as the backbone. We jointly estimate the location and uncertainty associated with the landmarks.

1. Uncertainty estimated by Cholesky Estimator Network (CEN).

- Output Cholesky coefficients L_{ij}, used to compute the covariance matrix.

2. Landmark location estimated by spatial mean of the ReLUed heatmap (H_{ij})

- Take ReLU over the heatmap to select the positive entries, then take the spatial mean.

3. Joint estimation of location and uncertainty using Gaussian Log Likelihood Loss

5. Experiments and Results on 300-W and Menpo

Experiment Setup:

- Split 1: Train = 3148 images of 300-W; Test = 689 images of 300-W
- Split 2: Train = 3837 images of 300-W; Test = 600 images of 300-W (Indoor/Outdoor) or 6679 frontal images of Menpo
- Metrics: Normalized Mean Error (NME); Area Under Curve (AUC) of fraction correct vs. error threshold

Evaluation of Landmark Prediction:

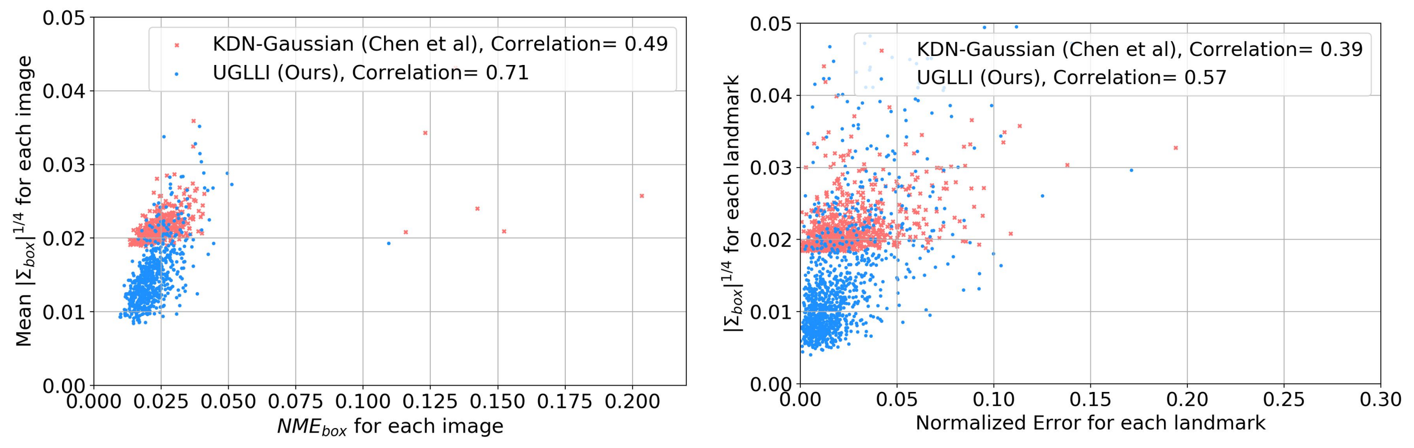
	Common	Challenge	Full
SAN [3]	3.34	6.60	3.98
DAN [4]	3.19	5.24	3.59
DU-Net [2] (public code)	2.97	5.53	3.47
UGLLI (Ours)	**2.87**	**5.08**	**3.23**

Table: $NME_{interocular}$ on 300-W (Split 1)

	NME_{box}(%) (↓)		AUC^7_{box}(%) (↑)	
	300-W	Menpo	300-W	Menpo
2D-FAN [5]	2.56	2.32	66.90	67.40
KDN-Gaussian [1]	2.49	2.26	67.30	68.40
UGLLI (Ours)	**2.24**	**2.20**	**68.27**	**69.85**

Table: NME and AUC on Split 2

Evaluation of Uncertainty Prediction:



6. Conclusion and Future Work

- Joint estimation of landmark location and uncertainty using UGLLI not only provides state-of-the-art uncertainty measures but also yields state-of-the-art estimates for the facial landmark locations.
- Future work includes application of this framework to other landmark regression problems, such as human body 2D pose estimation, and using estimated uncertainties to selectively improve the predictions.

References:

[1] Chen et al, Kernel density network for quantifying regression uncertainty in face alignment, NeurIPS Workshops 2018
[2] Tang et al, Towards Efficient U-Nets: A Coupled and Quantized Approach, ECCV 2018, TPAMI 2019
[3] Dong et al, Style aggregated network for facial landmark detection, CVPR 2018
[4] Kowalski et al, A convolutional neural network for robust face alignment. In CVPR Workshops, 2017
[5] Boulat et al, How far are we from solving the 2D & 3D face alignment problem?, ICCV 2017