A Multi-Task Bayesian Deep Neural Net for Detecting Life-Threatening Infant Incidents From Head Images

Julius Wang, Jorma Laaksonen, Yi-Ping Liao, Bo-Zong Wu, and Shih-Yun Shen
Aalto University, Espoo, Finland ■ Independent Researcher ◆ Yun yun AI Baby camera Co., Ltd., Taiwan (R.O.C)

ICIP 2019

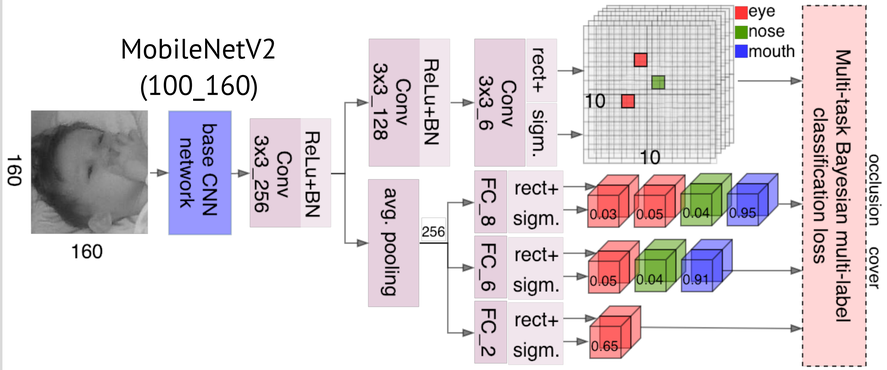
CuboAi

The Problem and Contributions

* **Analyze visual signs** that links to **discomforts** in infants
* **Detect**
  * Visible facial landmarks
  * Occlusions / Covering on facial parts
  * Eye openness

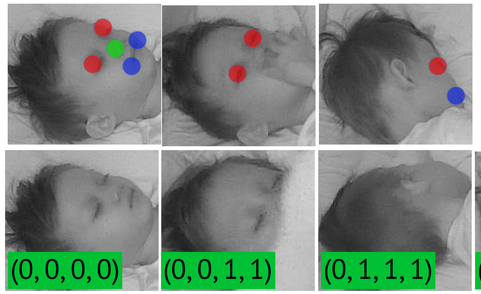
* Main Contributions
  * Presented a **novel computer vision application** for **neonatal care**
  * Proposed a **multi-task multi-label loss** that accounts for **(aleatoric) data uncertainty**
  * Showed that the model which **accounts for data uncertainty** outperforms other baseline models that do not

Network Architecture



YunInfants Dataset and Labeling Protocols

* We introduced YunInfants dataset, 19716 images captured by Cubo
  * Contain daytime and night-vision images, varying head positions
  * Age group from 0 to 2



Landmarks
* centers of eyes
* tip of the noise
* two mouth corners

Occlusions (1/0)
* eye, eyes
* nose
* mouth corners

Covering (1/0)
* eye
* nose
* mouth

Eye status

open: 1, not open: 0

Number of images in train / val / test: 12850 / 3211 / 3655

	train / val / test			
% of occl.	eye	eyes	nose	mouth
	37 / 38 / 38	21 / 21 / 15	38 / 38 / 32	34 / 35 / 39
% of cover.	eye	nose	mouth	
	12 / 13 / 13	21 / 21 / 24	34 / 35 / 39	
% of eye open	22 / 19 / 20			

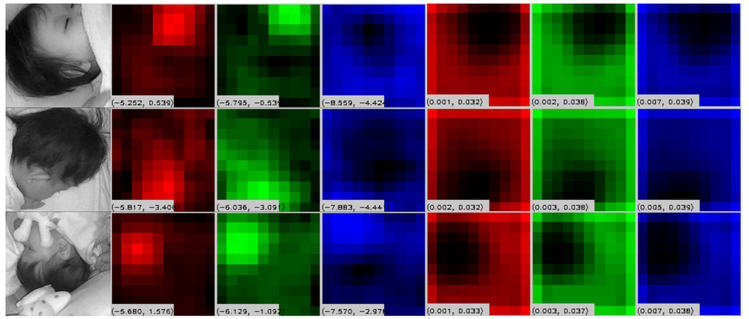
Multi-task Loss Functions with Aleatoric Uncertainty

Total loss: $L = L_{occ} + L_{cov} + L_{eye} + \alpha_{lm} L_{lm}$

Per-task loss: $L_* = -\frac{1}{T}\sum_{t=1}^{T}\frac{1}{c_*}\sum_{c=1}^{c_*}\log(\hat{p}^t_{*,c,i}),\ i$: target class

c_*: number of attributes to predict for task $* = \{occ, cov, eye, lm\}$
$\hat{p}^t_{*,c,i}$: prob. of c^{th} attribute estimated from the t^{th} sample drawn from $\mathcal{N}(f_{*,c,i}, \sigma^2_{*,c,i})$

Prediction and Uncertainty Visualizations



* Landmark detection provides **critical visual cues** for other tasks and can be used as **model's diagnosis tool**
* **Much higher uncertainty** revealed in the **background pixels**
* Uncertainty map acts as a face detector

Experimental Results

	occlusion					cover				eye	all tasks
F1-score (%)	eye	eyes	nose	mouth	avg.	eye	nose	mouth	avg.	openness	avg.
{o,c,e}+baux	80.71	75.64	83.24	89.97	82.39	69.52	78.01	81.11	76.21	82.56	80.39
{o,c,e}+baux+bayes	80.01	72.06	84.09	89.63	81.45	70.78	77.12	80.91	76.27	83.20	80.31
o+c+e+bayes	79.96	76.76	83.93	90.14	82.70	72.71	78.72	81.50	77.64	83.29	81.21
o+c+e+baux	80.26	75.02	83.56	90.00	82.23	71.17	79.00	81.76	77.31	**84.51**	81.35
o+c+e+baux+bayes	**82.19**	**77.16**	**85.64**	**91.09**	**84.02**	**74.82**	**79.57**	**83.11**	**79.17**	83.92	**82.37**

Aalto University
School of Science

{tzu-jui.wang, jorma.laaksonen}@aalto.fi, yiping.ncu@gmail, {barry, steven}@yunyun.cloud