BAYESIAN CROWD COUNTING

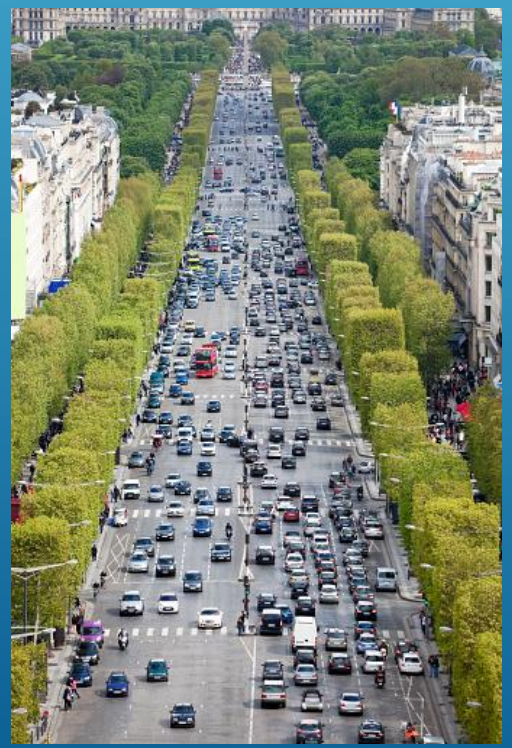
Original paper:

Bayesian Loss for Crowd Count Estimation with Point Supervision

Zhiheng Ma, Xing Wei, Xiaopeng Hong, Yihong Gong

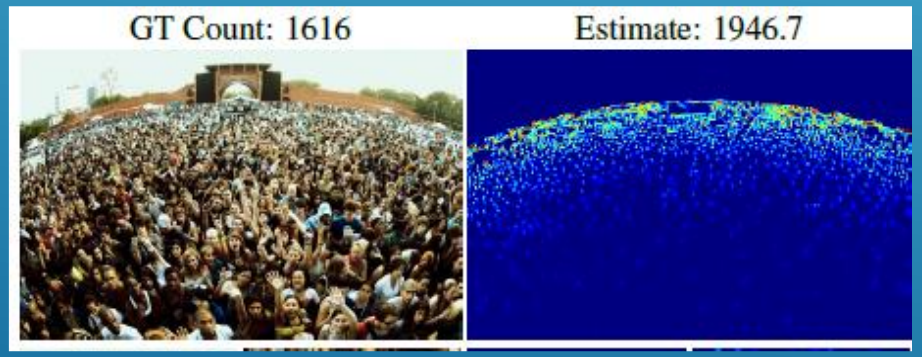
TABLE OF CONTENTS

- Introduction
- Density map estimation approach and its limits
- Bayesian/Bayesian+ loss approach
- Implementation details of the original model
- Extension of the training on the JHU-dataset
- Demo



INTRODUCTION

- **Task**: count of all the occurences of an object in an image or a video frame
- **Applications**: count of participants in social or sport events, count of cars in traffic congestions, count of cells and bacteria from microscopic images, etc.
- **Problems**:
  1) Dense crowds often have heavy overlaps and occlusions between each other;
  2) Perspective effects may cause large variations in human size, shape, and appearance in the image.



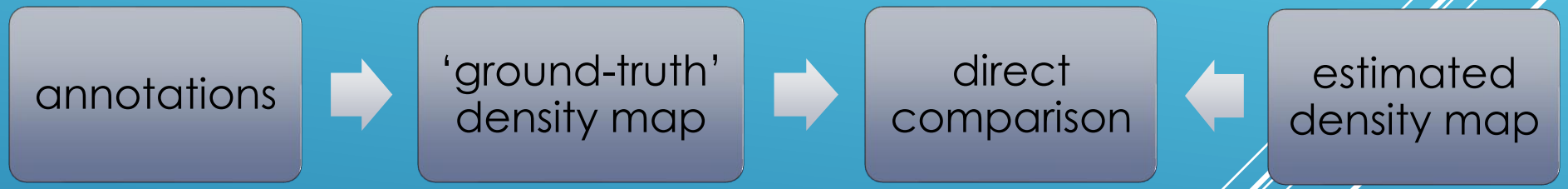
DENSITY MAP ESTIMATION APPROACH

- Datasets for training crowd counting estimators only provide **point annotations for each training image**, i.e., only one pixel of each person is labeled;
- The most common approach for using these annotations is to first convert the point annotations into a **"groundtruth" density map** $\boldsymbol{D}$, obtained by filtering through a Gaussian kernel;
- **Loss function**:

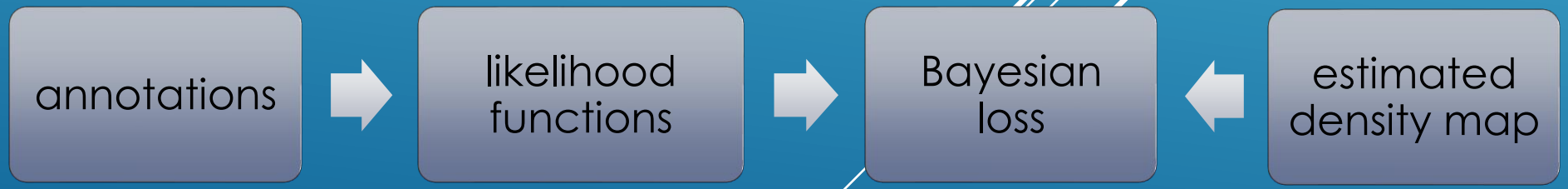
$$\mathcal{L} = \sum_{m=1}^{M} \mathcal{F}\left(D^{gt}(x_m) - D^{est}(x_m) \right) .$$

BAYESIAN LOSS APPROACH

- **Baseline approach**:



- **Bayesian Loss approach**:



BAYESIAN LOSS APPROACH

Likelihood function with the same form of the Gaussian kernel:

$$p(\boldsymbol{x} = \boldsymbol{x}_m | y = y_n) = \mathcal{N}(\boldsymbol{x}_m; \boldsymbol{z}_n, \sigma^2 \mathbb{1}_{2\times 2}),$$

$$\begin{cases} \boldsymbol{x}_m \rightarrow \text{spatial coordinates} \\ y_n \rightarrow \text{labels} \\ \boldsymbol{z}_n \rightarrow \text{annotated point coordinates} \end{cases}$$

Bayes' theorem:

$$p(y_n|\boldsymbol{x}_m) = \frac{p(\boldsymbol{x}_m|y_n)p(y_n)}{p(\boldsymbol{x}_m)}$$

$$\mathcal{L}_{Bayes} = \sum_{n=1}^{N} \mathcal{F}(1 - \boldsymbol{E}[c_n]) \text{ , with}$$

$$\boldsymbol{E}[c_n] = \sum_{m=1}^{M} p(y_n|\boldsymbol{x}_m)\, \boldsymbol{D}^{est}(\boldsymbol{x}_m)$$

BAYESIAN+ LOSS APPROACH

For background pixels that are far away from any of the annotation points, it makes no sense to assign them to any head label.

An additional background label $y_0 = 0$ is introduced

the Gaussian kernel is used to build the background likelihood after defining

$$z_0^m = z_n^m + d\frac{x_m - z_n^m}{\|x_m - z_n^m\|^2}$$

With this approach, pixels from the density map far away from any annotation will be assigned to the background

IMPLEMENTATION DETAILS

- **NETWORK**

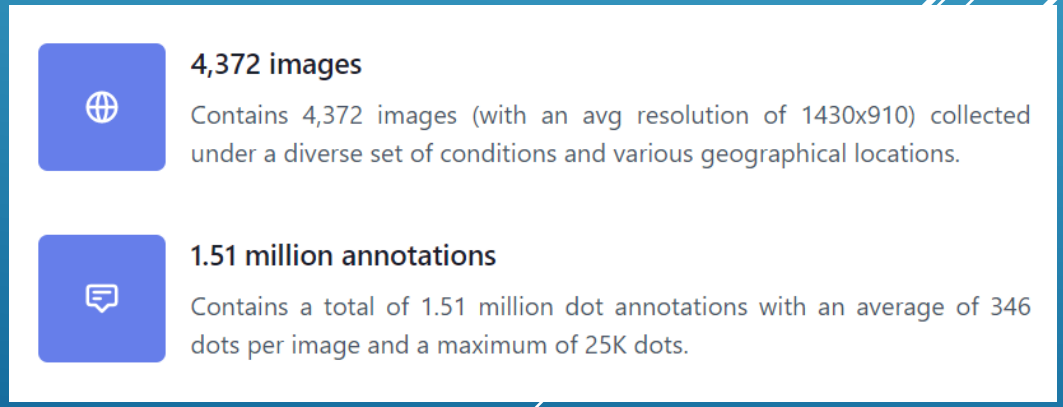
The model uses a VGG19 backbone pretrained on Imagenet and a regression layer that produces the density map.

- **PERFORMANCES OF THE MODEL**

Datasets	UCF-QNRF		ShanghaiTechA		ShanghaiTechB		UCF_CC_50	
Methods	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
CROWD-CNN [53]	-	-	181.8	277.7	32.0	49.8	467.0	498.5
MCNN [57]	277	426	110.2	173.2	26.4	41.3	377.6	509.1
CMTL [40]	252	514	101.3	152.4	20.0	31.1	322.8	341.4
SWITCH-CNN [3]	228	445	90.4	135.0	21.6	33.4	318.1	439.2
CP-CNN [41]	-	-	73.6	106.4	20.1	30.1	295.8	320.9
ACSCP [37]	-	-	75.7	102.7	17.2	27.4	291.0	404.6
D-CONVNET [38]	-	-	73.5	112.3	18.7	26.0	288.4	404.7
IG-CNN [2]	-	-	72.5	118.2	13.6	21.1	291.4	349.4
IC-CNN [32]	-	-	68.5	116.2	10.7	16.0	260.9	365.5
SANET [4]	-	-	67.0	104.5	8.4	13.6	258.4	334.9
CL-CNN [16]	132	191	-	-	-	-	-	-
BASELINE	106.8	183.7	68.6	110.1	8.5	13.9	251.6	331.3
Our BAYESIAN	92.9	163.0	64.5	104.0	7.9	13.3	237.7	320.8
Our BAYESIAN+	**88.7**	**154.8**	**62.8**	**101.8**	**7.7**	**12.7**	**229.3**	**308.2**

OUR APPROACH

We extended the original training from the weights obtained by the authors on the SHANGHAI-A dataset using one closer to the images required to be processed. This required modifying the preprocessing code to adapt the program to .txt labels, in contrast to the original .mat format. We chose to use a subset of 1000 random samples from the JHU-CROWD++ dataset.



DEMO

https://i.imgur.com/ZHqkf7N.gif