

Crowd Counting: My project

Computer Vision

Mattia Paolacci

The goal of this project

- The project has aimed to try to reproduce what that has been described into the paper “Towards Perspective-Free Object Counting with Deep Learning” just starting from the knowledge providing by the such paper.
- I try to build the CCNN model and another model taken from [github](#) by using the Tensorflow framework with the Keras API.
- Then, given a dataset of images I build the related density maps then I proceeded by training the CCNN model, at last I observed the results from a qualitative point of view of both models.

The pipeline

- Build the density maps
- Extract the patches from the images
- Training of the CCNN and the *other* model in Colab by using the GPUs
- Evaluation of the two models
- A qualitative and quantitative comparison of the density maps predicted by the two models.

How the density map was built (1)

```
# Build the empty density map
density_map = np.zeros((height, width), dtype=np.float32)
n_points = len(head_points)

if(n_points <= 0):
    return density_map.tolist()

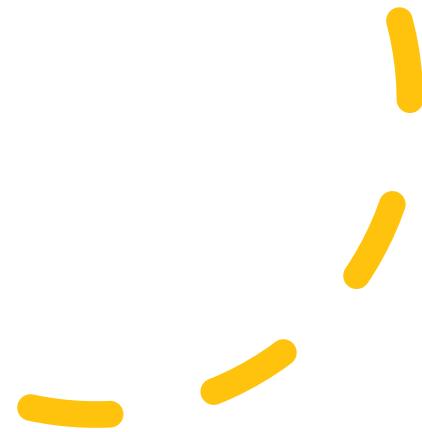
for p in head_points:
    dots = np.zeros((height, width), dtype=np.float32)
    dots[p[1], p[0]] = 1.
    density_map += gaussian_filter(dots, sigma=SIGMA, mode='constant')
```

- Taking up the definition:
“Given an image I , the density map D_I is defined as a sum of *Gaussian functions* centered in each dot annotation”.
- I used the `gaussian_filter` from *ndimage* module of *scipy*.



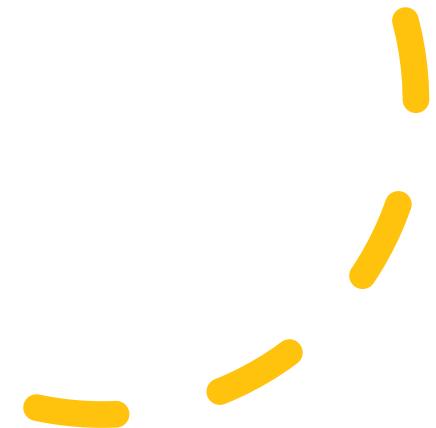
How the density map was built (2)

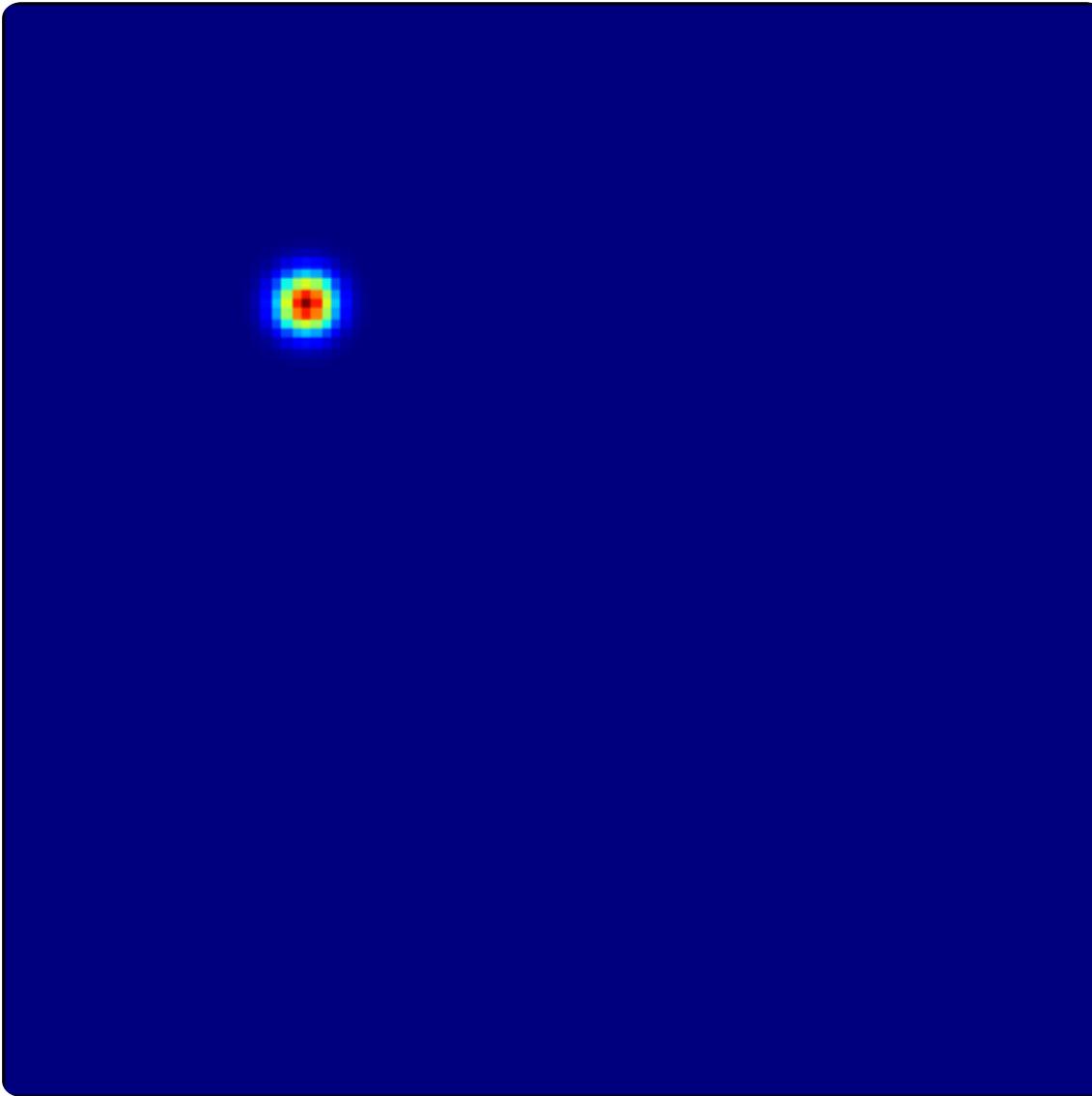
What happens from a visual
point of view?



How the density map was built (3)

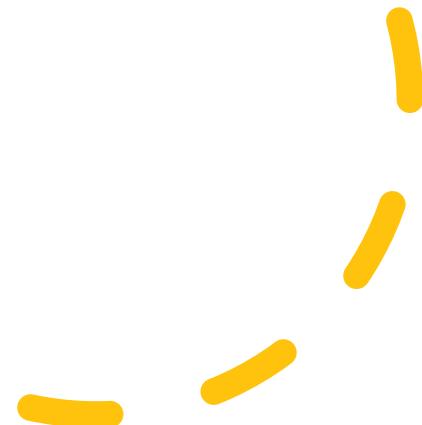
What happens from a visual
point of view?





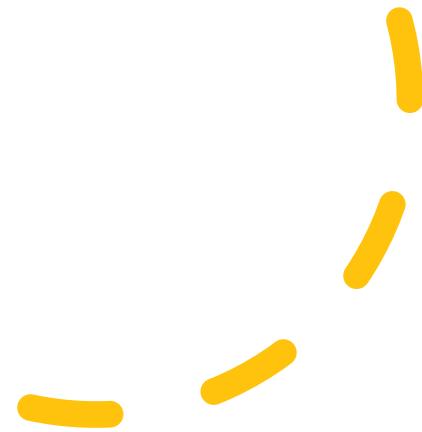
How the density map
was built (4)

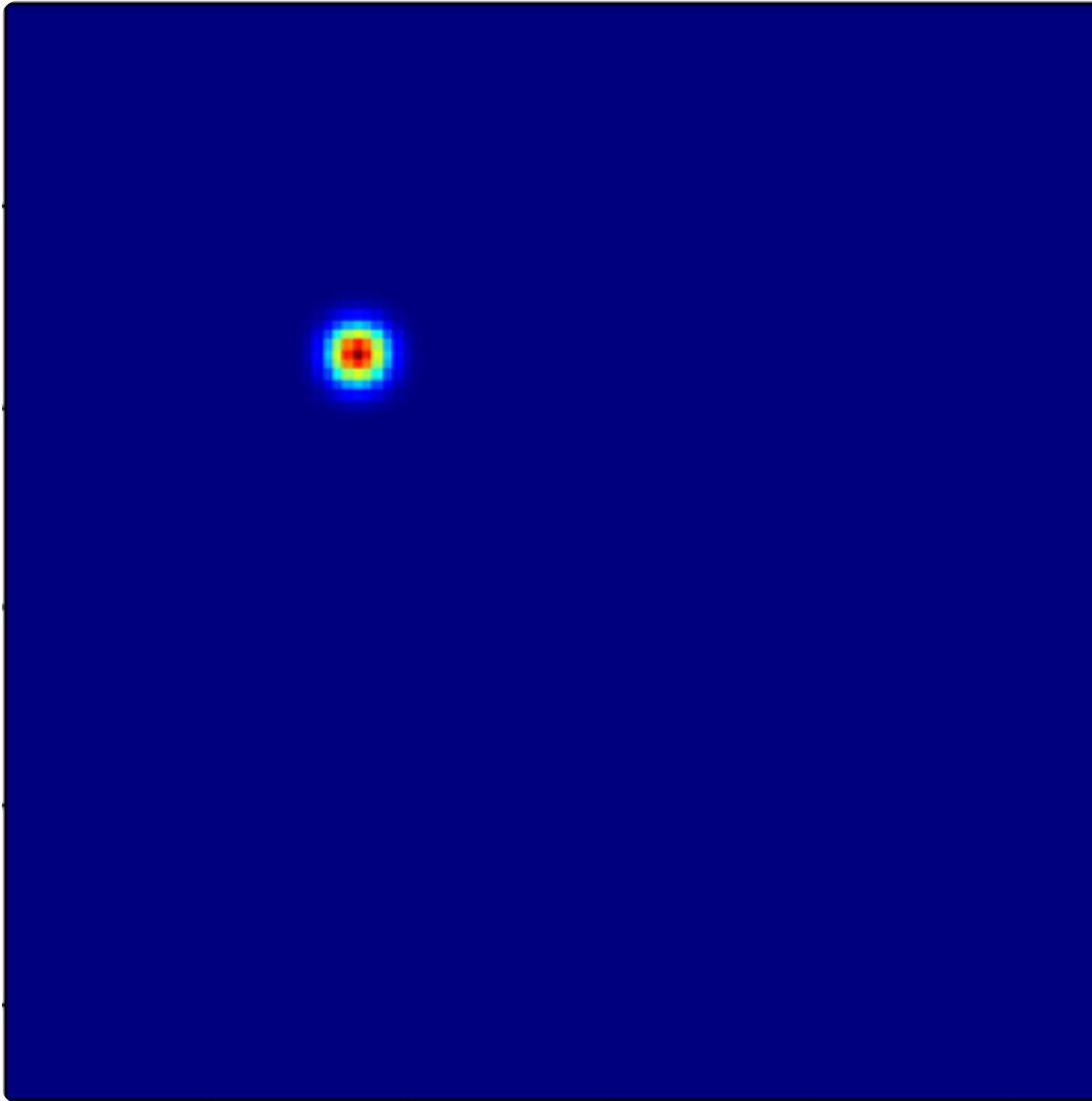
What happens from a visual
point of view?



How the density map was built (5)

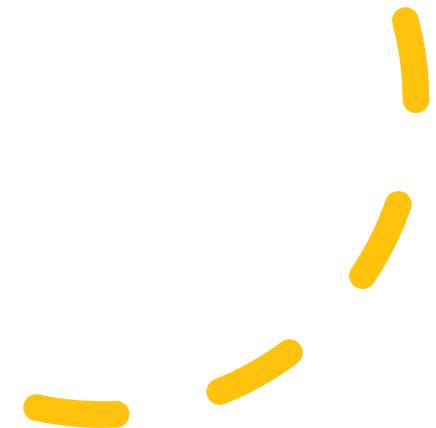
What happens from a visual
point of view?

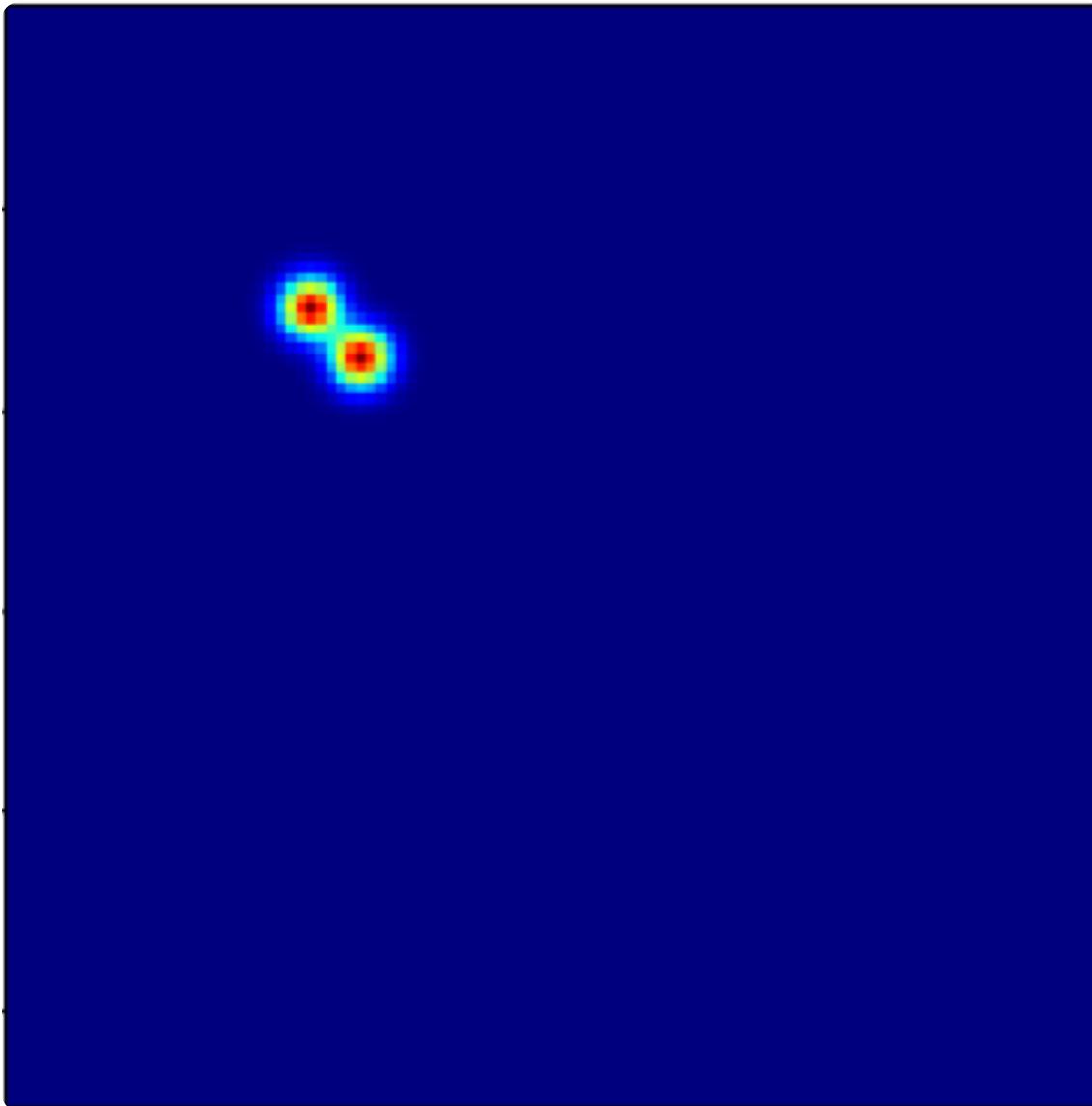




How the density map was built (6)

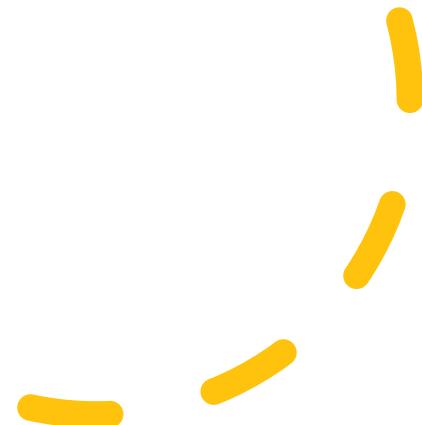
What happens from a visual point of view?





How the density map
was built (7)

What happens from a visual
point of view?



Some considerations of the construction of the density maps

- This procedure takes very long time with many big images (e.g., 2k, 4k or more). Therefore, I scaled the images before proceeding further and I implemented a multithread algorithm so that reduce the time taken by the construction of a single density map.
- The *sigma* value (that is the standard deviation) has been chosen in an empirically way by considering that the more sigma grows the more the gaussian function of a single point gets smoothed. Note that the more *sigma* is high the more the time *gaussian_filter* function needs.

A quantitative comparison between two models (1)

Both following results have been produced by a training of 100 epochs and a learning rate of 0.001

- batch size=128

MODEL	MSE	MAE	VAL_MSE	VAL_MAE	TEST_MSE	TEST_MAE
CCNN	4227.1	28.7	12945.1	40.7	5806.8	35.1
Jplumail_model	7679.8	48.2	5275.2	40.2	6926.0	45.1

- batch size=64

MODEL	MSE	MAE	VAL_MSE	VAL_MAE	TEST_MSE	TEST_MAE
CCNN	3165.2	27.8	3865.6	22.6	3704.8	26.0
Jplumail_model	6493.7	39.7	5128.5	39.7	5845.8	38.3

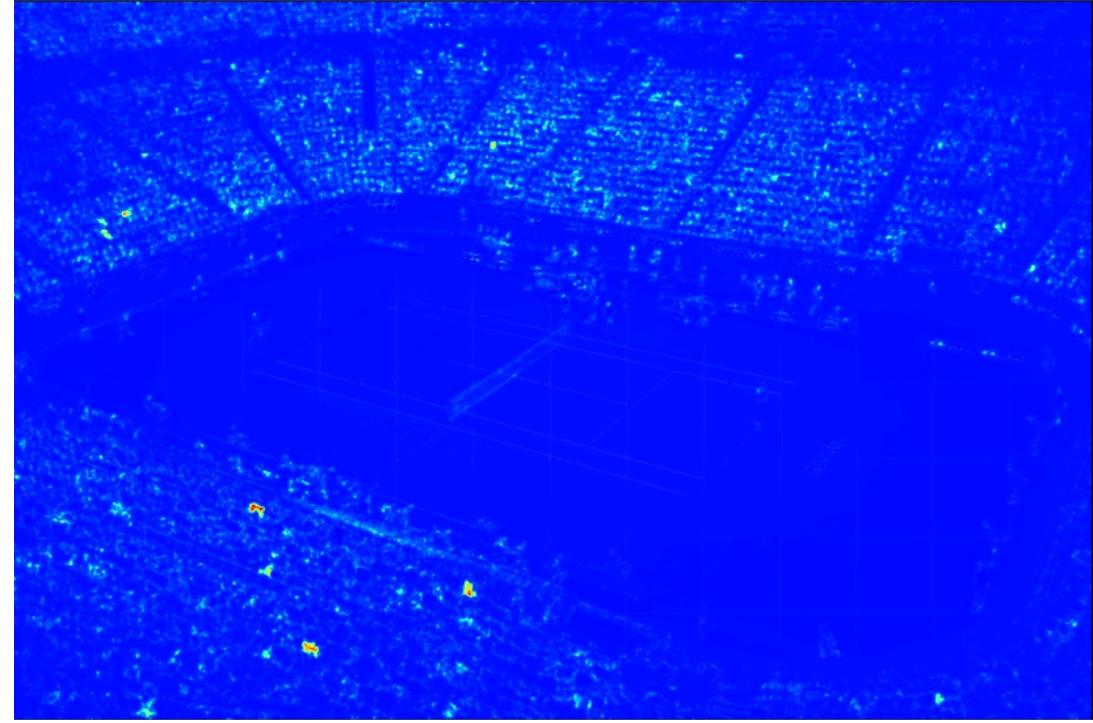
A quantitative comparison between two models (2)

Both following results have been produced by a training of 100 epochs and a learning rate of 0.001.

The result below are calculated on 100 entire images where the mean number of humans is 295, instead the max is 2925 and the min is 0.

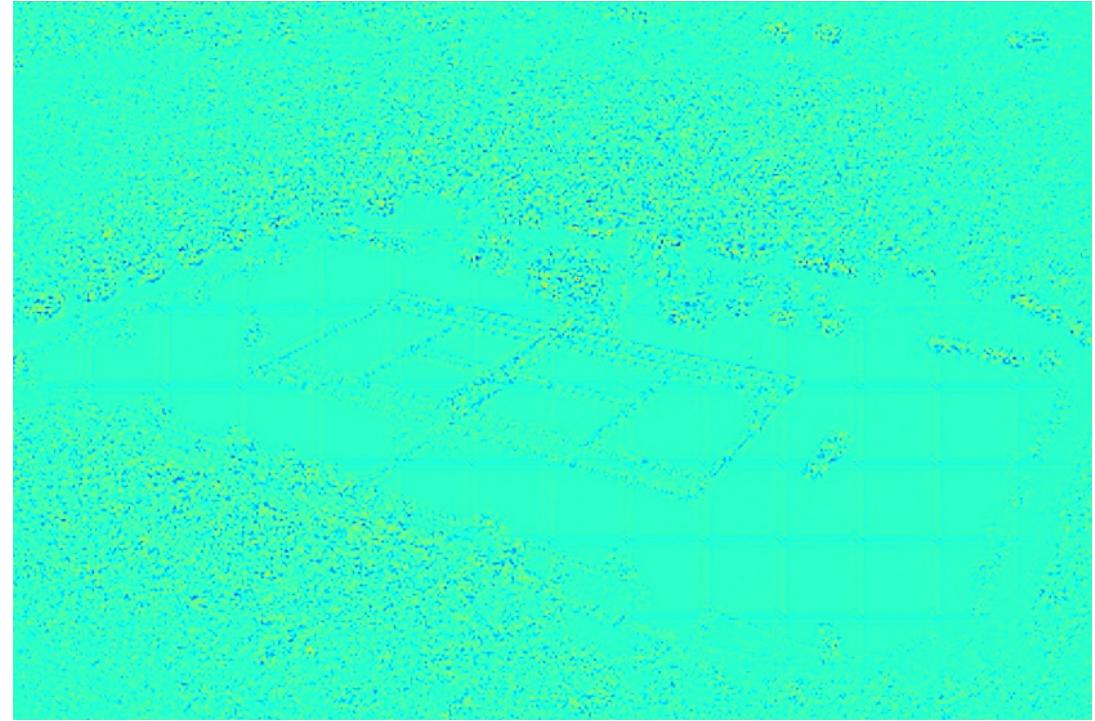
- batch size=64

MODEL	MAE
CCNN	1130.5
Jplumail_model	1913.7



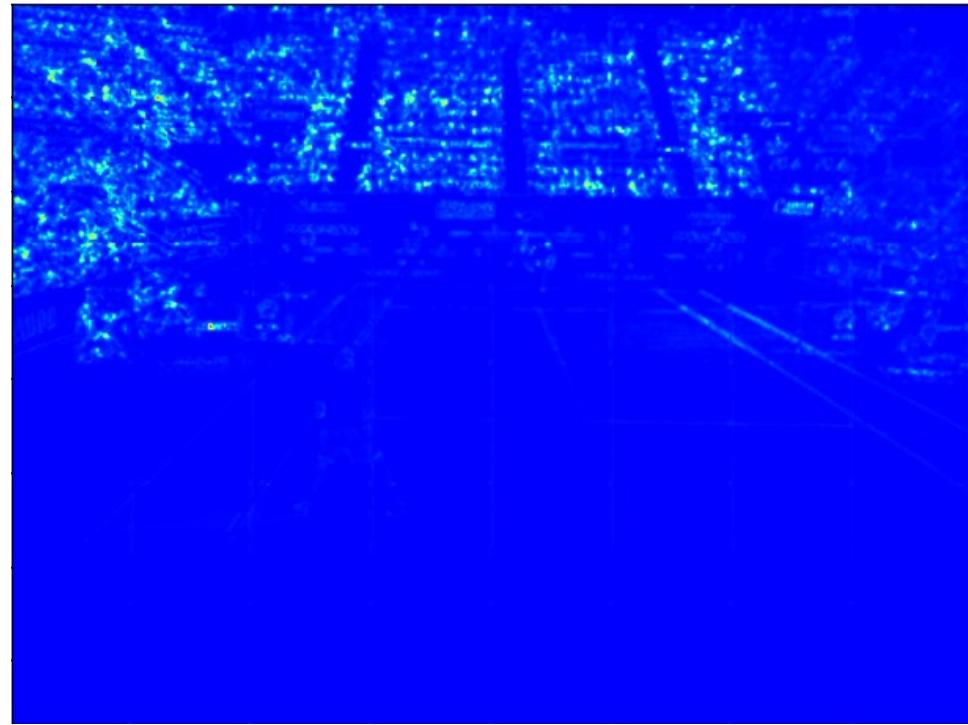
Qualitative results of the CCNN model (1)

- Ground-truth: 3081
- Predicted: 3148



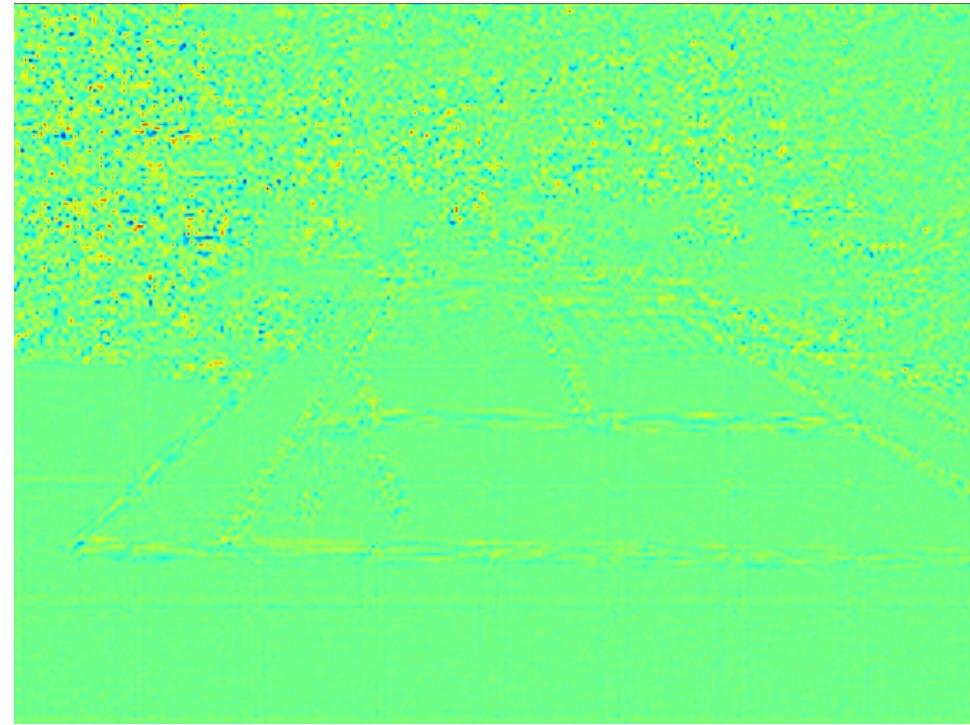
Qualitative results of the *other* model (1)

- Ground-truth: 3081
- Predicted: 7007



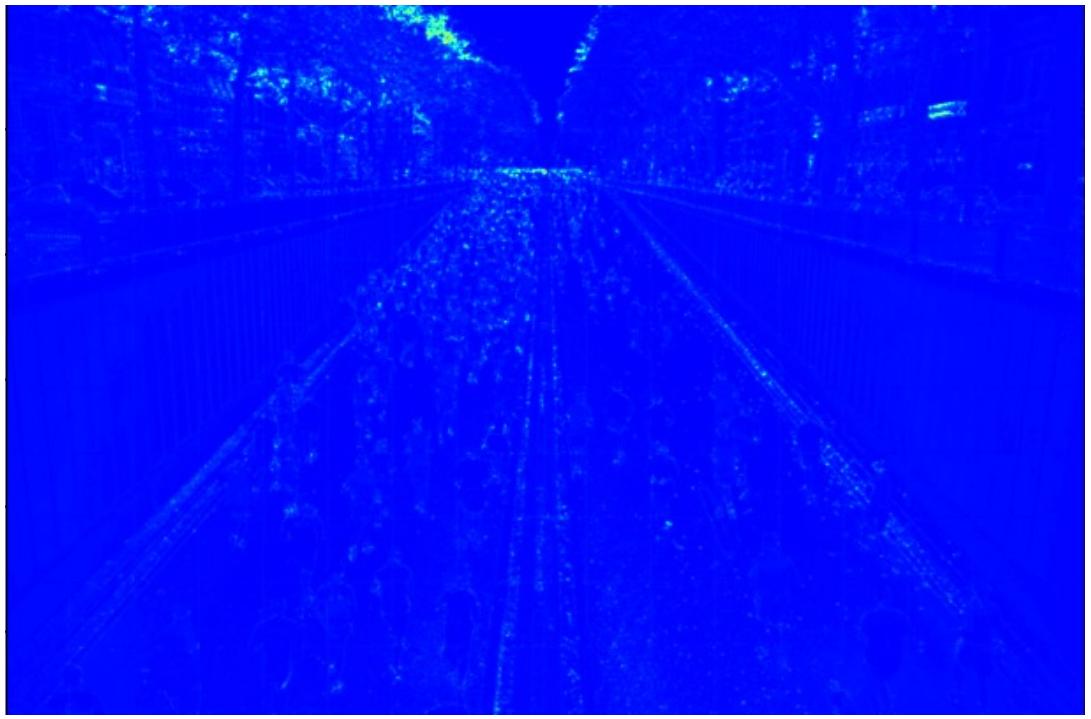
Qualitative results of
the CCNN model (2)

- Ground-truth: 637
- Predicted: 554



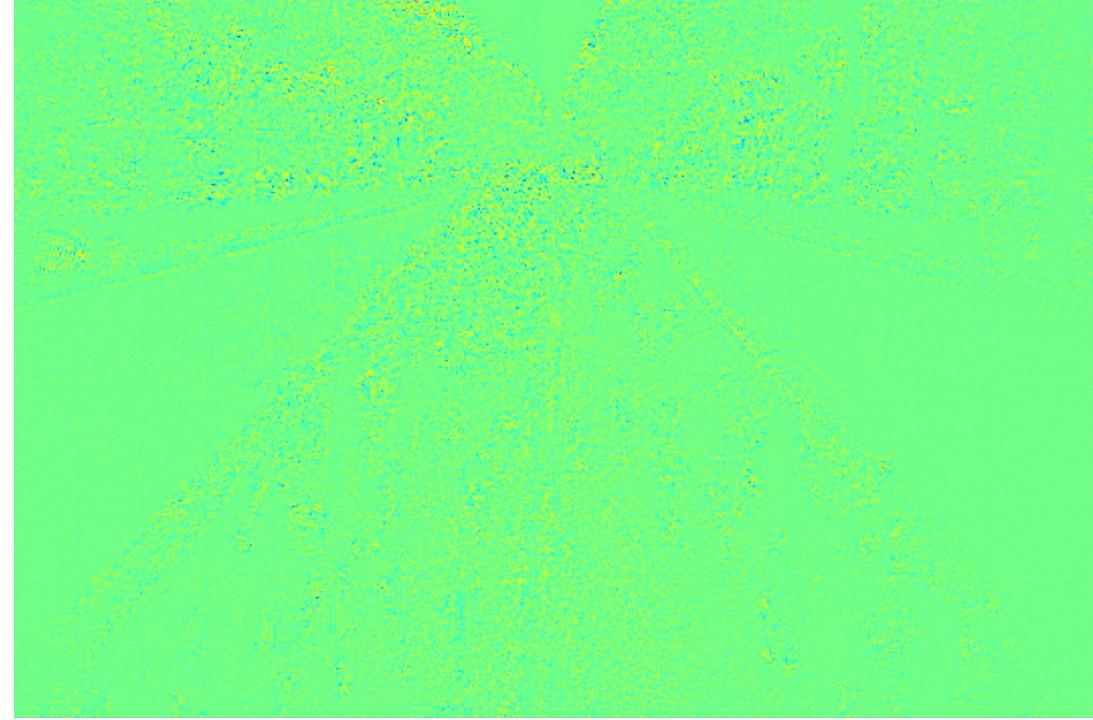
Qualitative results of
the *other* model (2)

- Ground-truth: 637
- Predicted: 192



Qualitative results of the CCNN model (3)

- Ground-truth: 286
- Predicted: 690



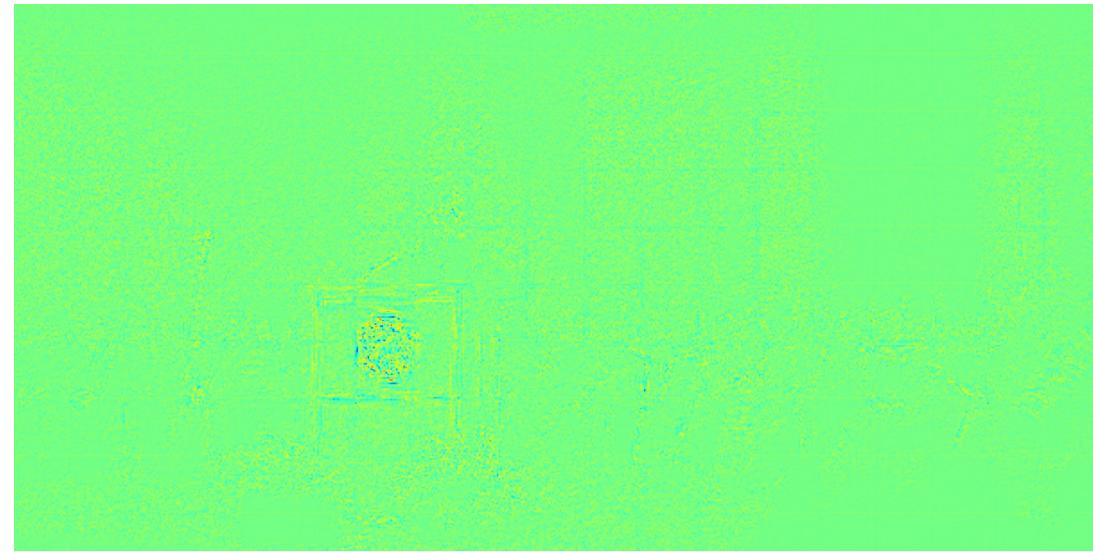
Qualitative results of
the *other* model (3)

- Ground-truth: 286
- Predicted: 7683



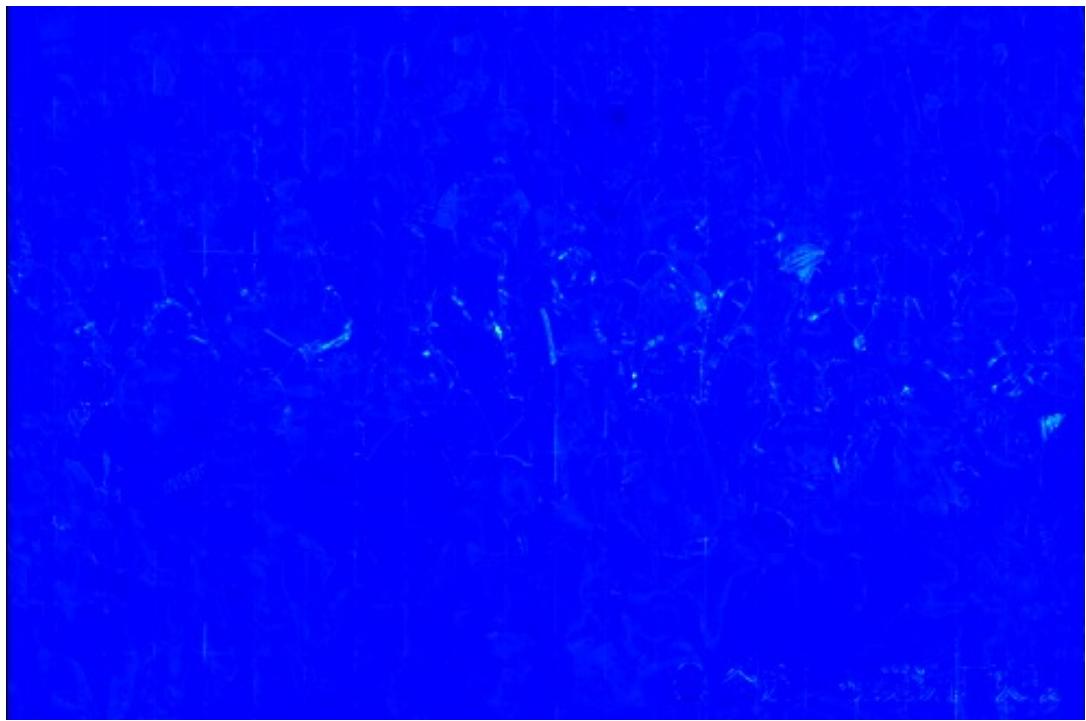
Qualitative results of the CCNN model (4)

- Ground-truth: 56
- Predicted: 169



Qualitative results of
the *other* model (4)

- Ground-truth: 56
- Predicted: 9144



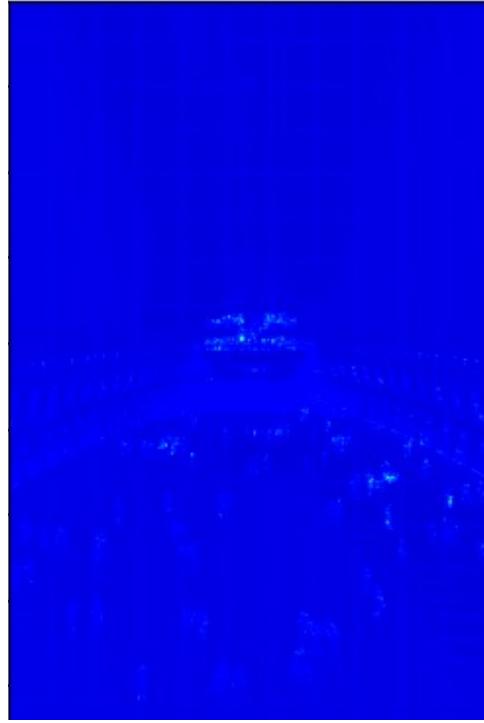
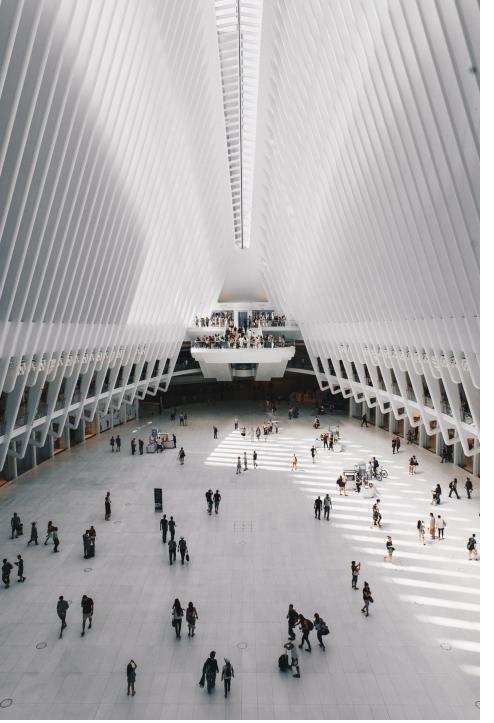
Qualitative results of the CCNN model (5)

- Ground-truth: 145
- Predicted: 387



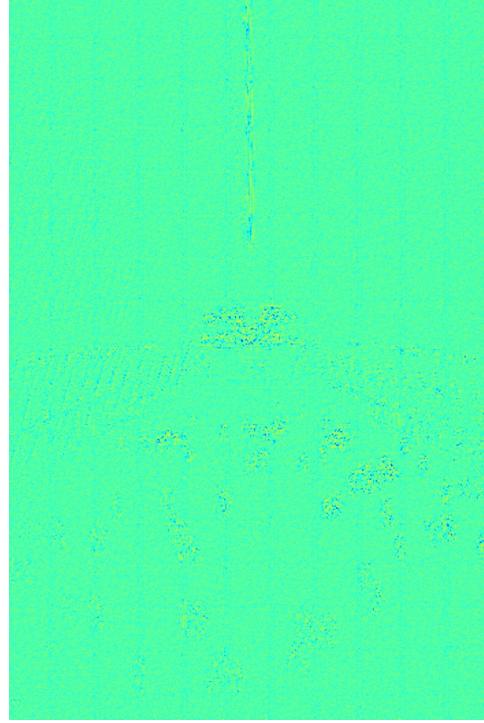
Qualitative results of the *other* model (5)

- Ground-truth: 145
- Predicted: 742



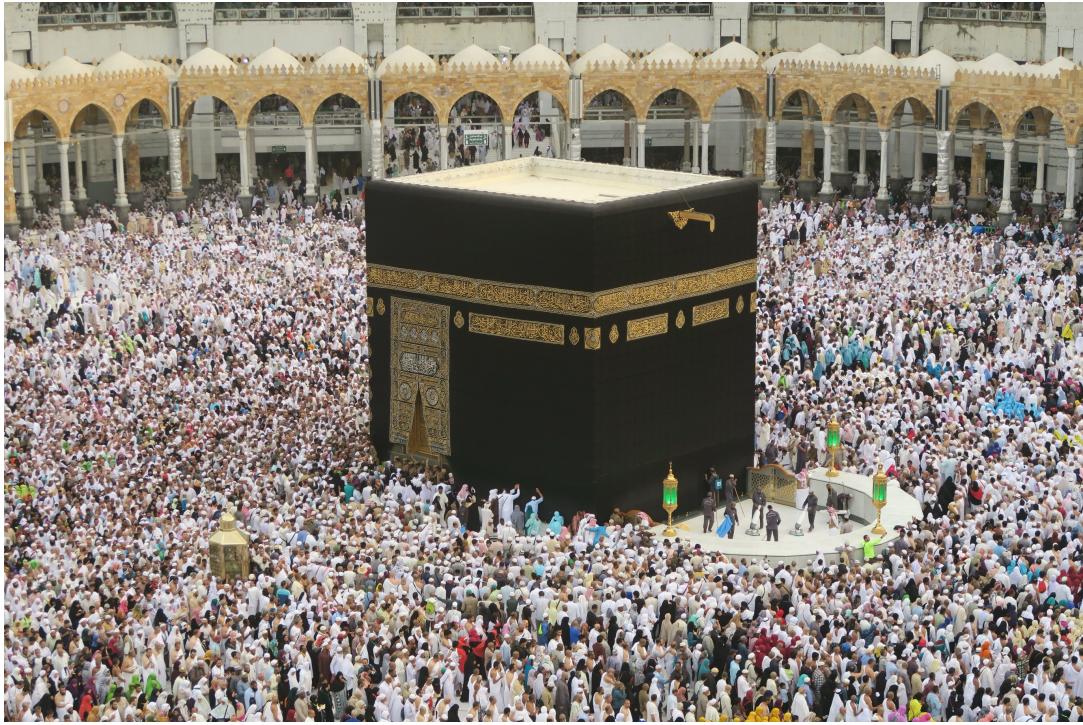
Qualitative results of the CCNN model (6)

- Ground-truth: 264
- Predicted: 311



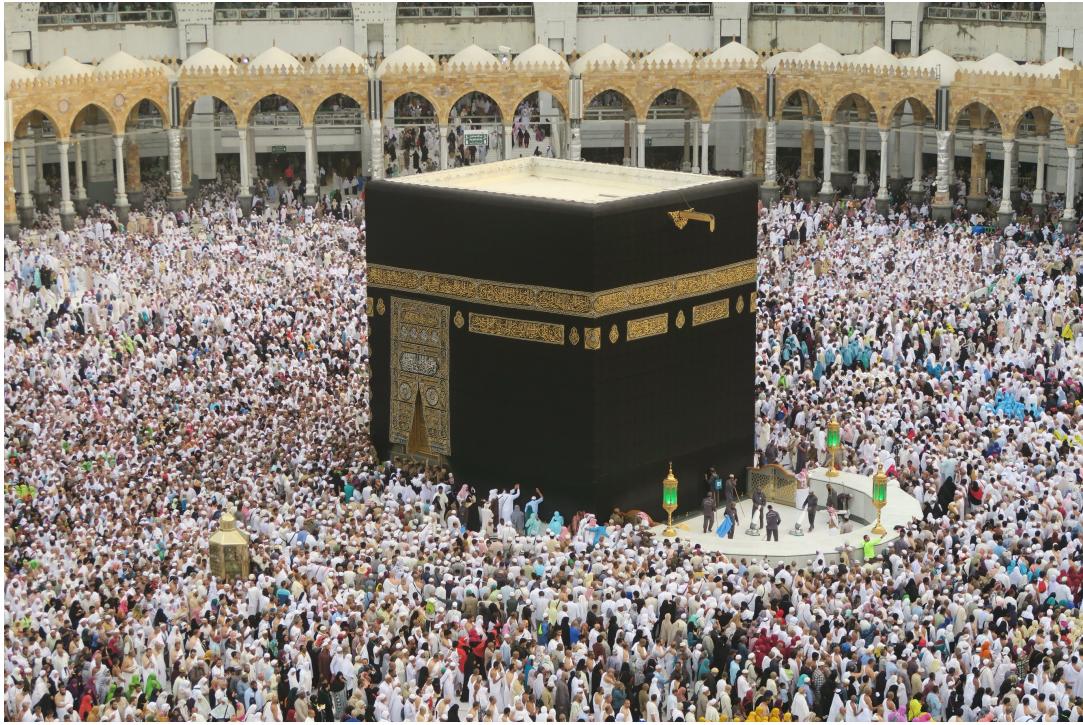
Qualitative results of the *other* model (6)

- Ground-truth: 264
- Predicted: 8815



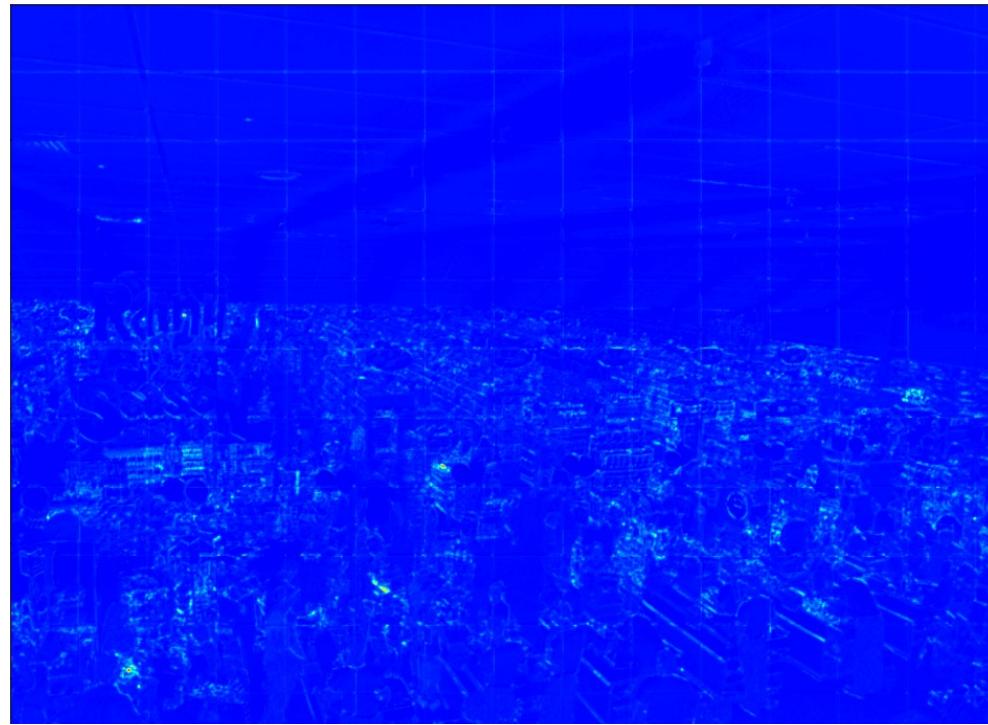
Qualitative results of
the CCNN model (7)

- Ground-truth: 6799
- Predicted: 1847



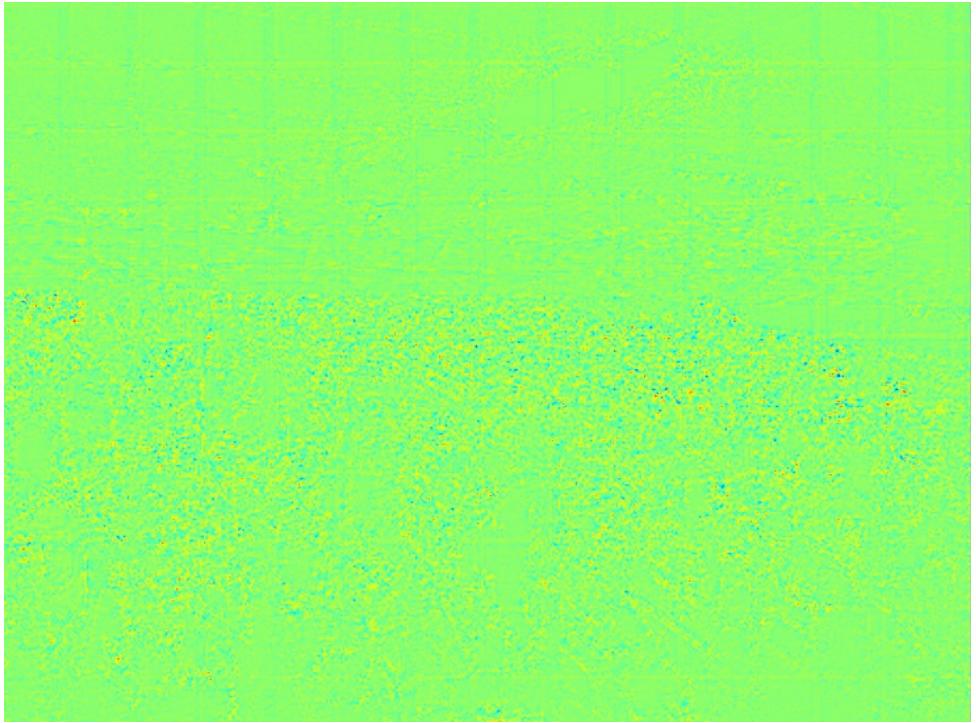
Qualitative results of
the *other* model (7)

- Ground-truth: 6799
- Predicted: 9538



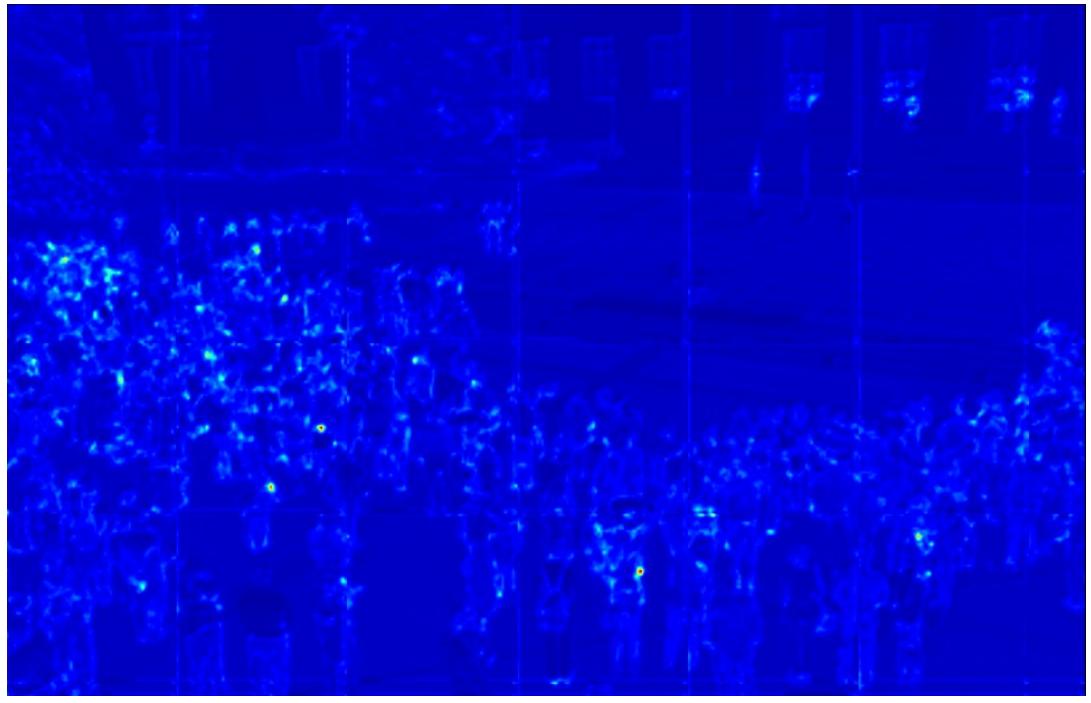
Qualitative results of
the CCNN model (8)

- Ground-truth: 104
- Predicted: 1449



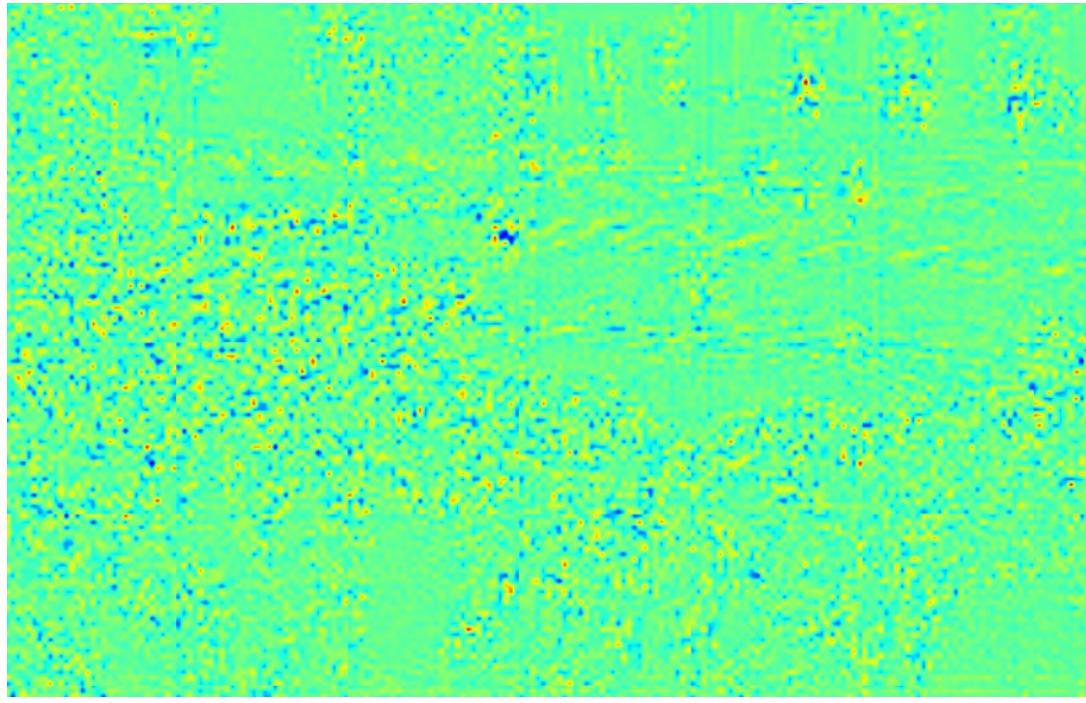
Qualitative results of
the *other* model (8)

- Ground-truth: 104
- Predicted: 1233



Qualitative results of the CCNN model (9)

- Ground-truth: 228
- Predicted: 187



Qualitative results of
the *other* model (9)

- Ground-truth: 228
- Predicted: 356



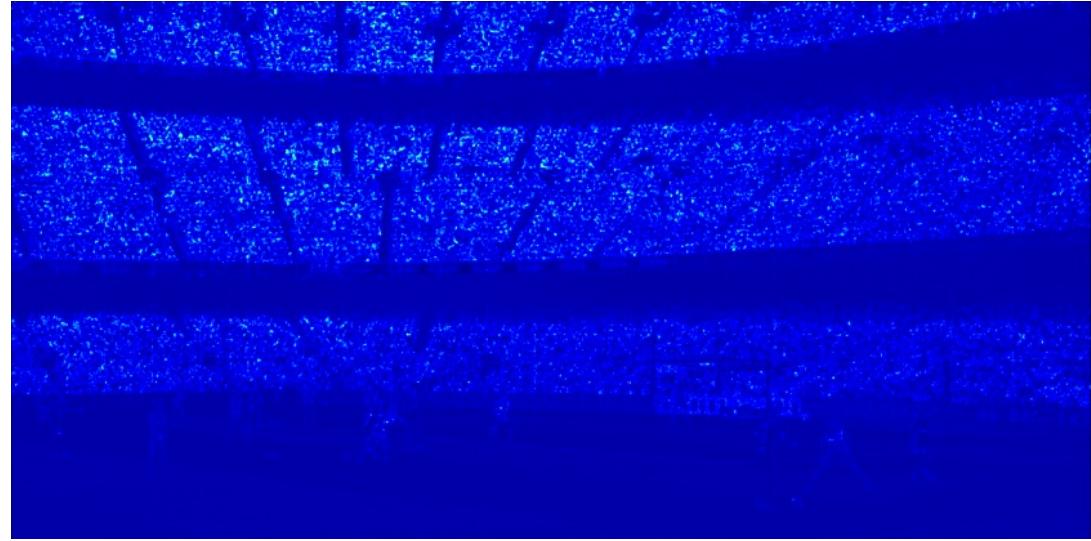
Qualitative results of the CCNN model (10)

- Ground-truth: 913
- Predicted: 21508



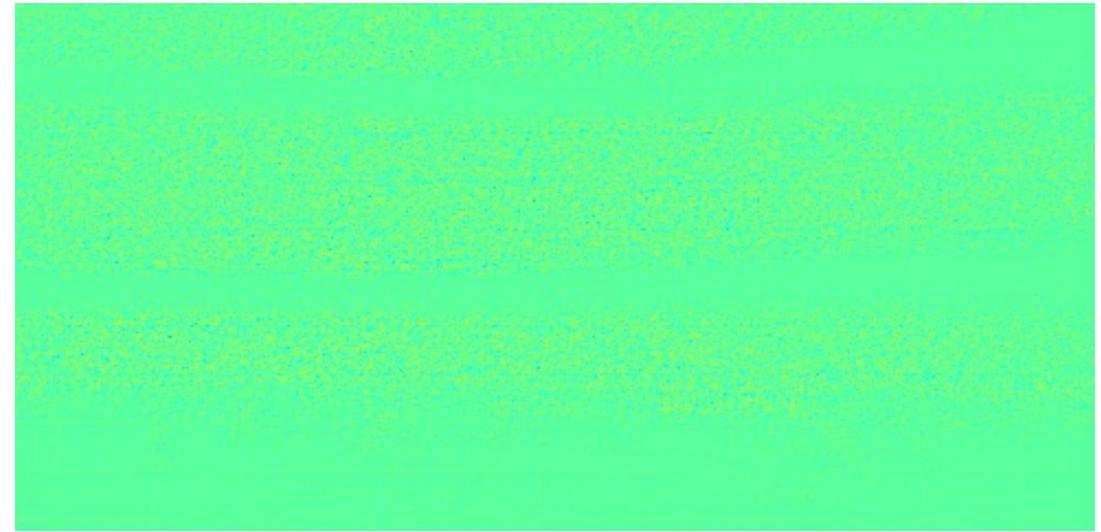
Qualitative results of the *other* model (10)

- Ground-truth: 913
- Predicted: 24042



Qualitative results of the CCNN model (10)

- Ground-truth: 8219
- Predicted: 11600



Qualitative results of
the *other* model (10)

- Ground-truth: 8219
- Predicted: 15496

Thanks for the attention ☺