Detecting Nonexistent Pedestrians

Jui-Ting Chien
Chia-Jung Chou
Hwann-Tzong Chen

Morning Session: Scene Understanding Workshop (SUNw'17)

Organizers: Bolei Zhou, Aditya Khosla, Jianxiong Xiao, James Hays

Afternoon Session: Large SUN Challenge (LSUN'17)

Organizers: Fisher Yu, Peter Kontschieder, Shuran Song, Ming Jiang, Yinda Zhang, Catherine Qi Zhao, Thomas Funkhouser, Jianxiong Xiao

LIP LOOK INTO PERSON

Login

HOME OVERVIEW LEADERBOARD BENCHMARKS SUBMIT CONTACT

Leaderboard for Our CVPR-2017 Workshop Challenge

The challenge is ended. We have received 63 submissions during that time period. Following is the final leader-board, where the parsing challenge is ranked by Mean IoU(%) and the pose challenge is ranked by PCK. We have omitted results without clear description.

umbo computer vision

Human Pose Challenge

Show 10 entries

Search:

Ranking	Method	PCK	Details	Submit Time
1	NTHU-Pose	87.400	Details	2017-06-02 03:06:07
2	Pyramid Stream Network (Multi-Model)	82.100	Details	2017-06-03 08:03:30
3	BUPTMM-POSE	80.200	Details	2017-06-04 14:53:20
4	Hybrid Pose Matchine	77.200	Details	2017-06-04 13:38:59

Showing 1 to 4 of 4 entries

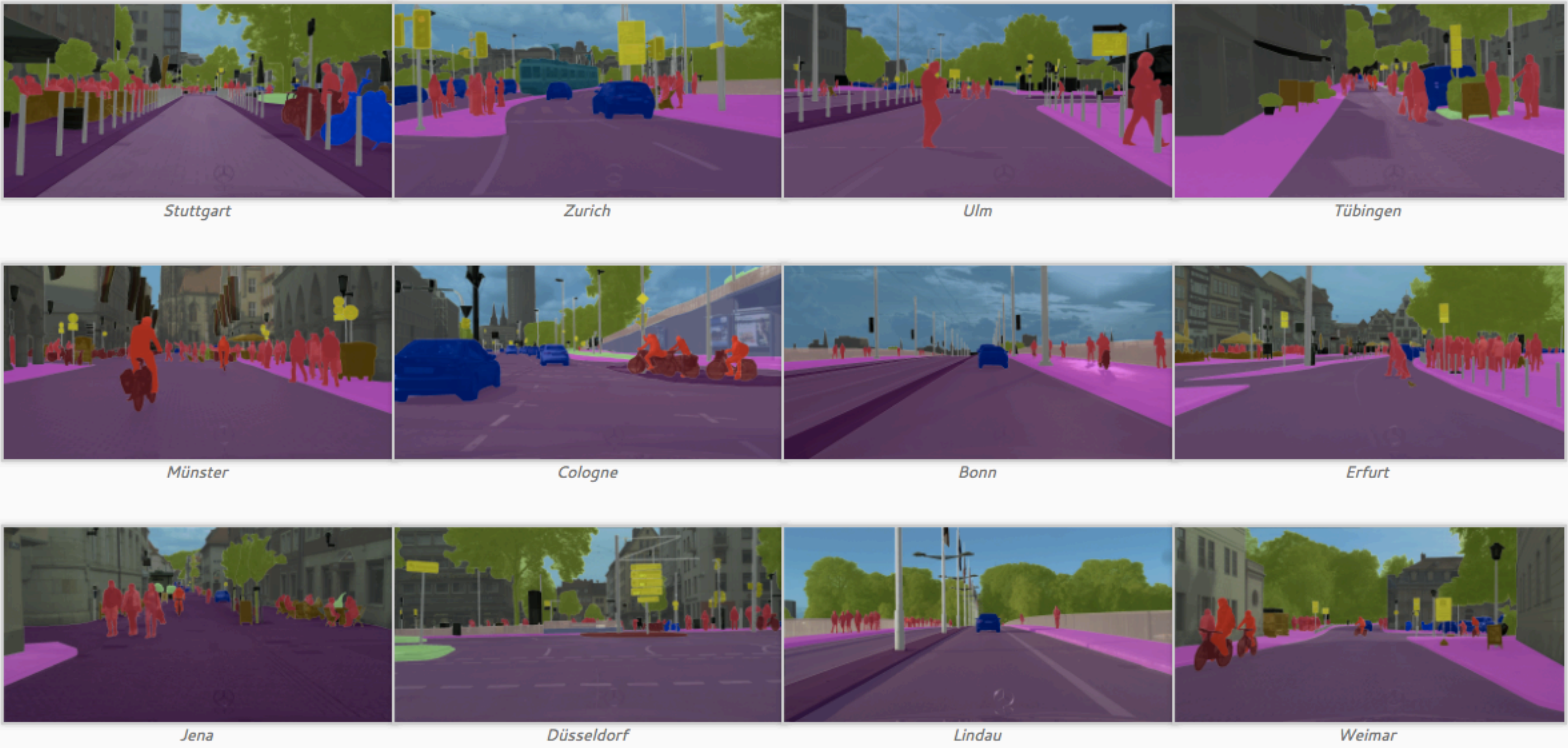
Previous 1 Next

Abbreviations

Detecting *nonexistent* pedestrians?

Urban Scene Understanding

What? Where? When? Why?



CITYSCAPES Dataset

Detecting Nonexistent Pedestrians vs. Detecting Pedestrians

- Predict from context
- Where to look at to find people?

Problem Definition

To predict the presence probabilities of nonexistent pedestrians in a street scene

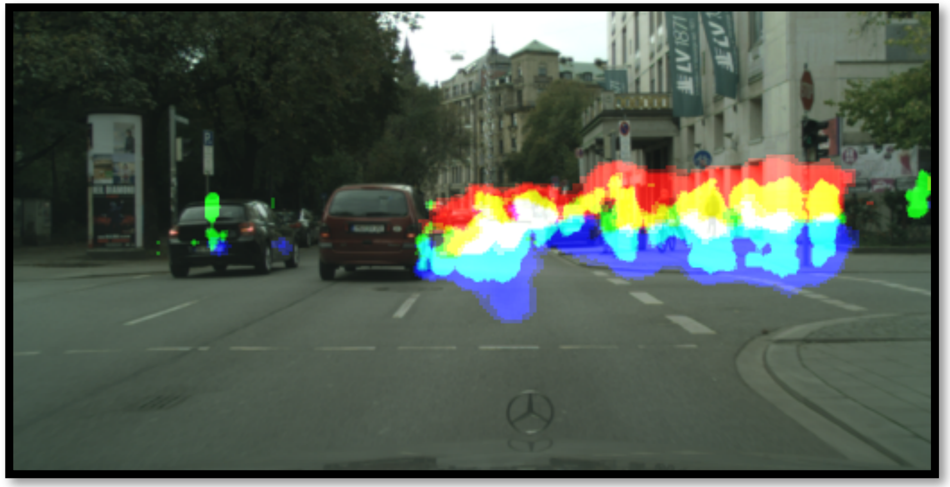
Input image

Probabilities map

Synthetic image

Experimental Results

a. Input image



b. Predicted heat map

c. Pedestrians placed arbitrary

d. Pedestrians placed according to (b)

Pipeline

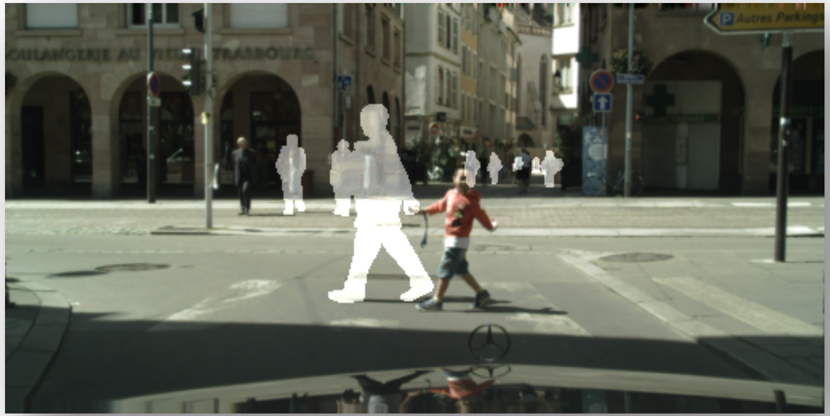
1. **Generate training data**
   - **Collection**
   - **Pose estimation**
   - **Inpainting**
   - **Input / output**

2. Train the network
   - Adversarial learning

3. Synthesize images
   - Pedestrian datasets
   - Synthesis

Training Data



Generate Training Data – Collection

Dataset

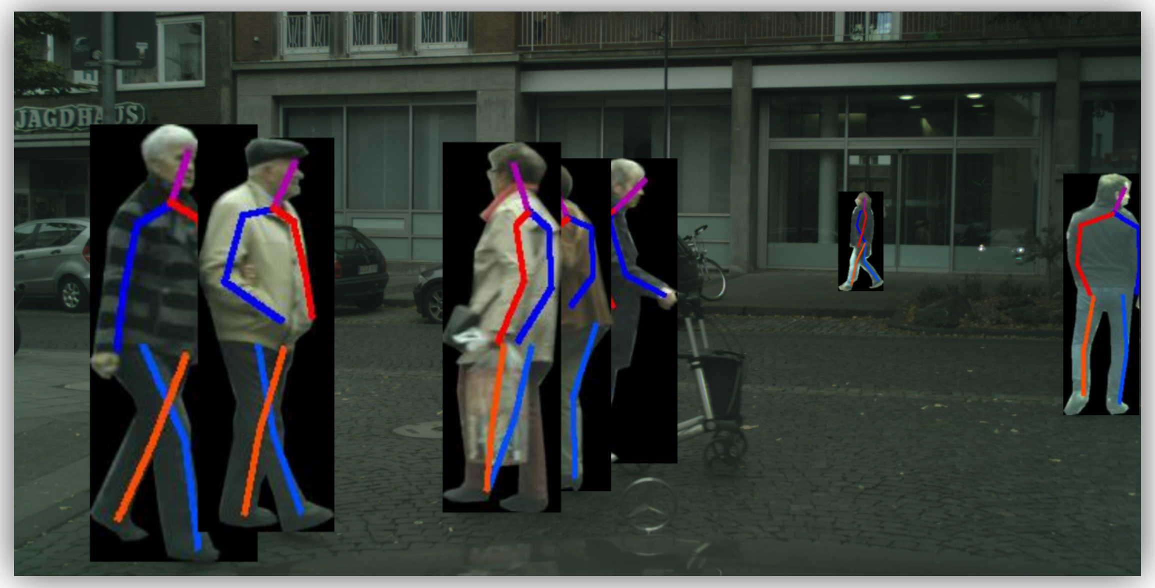
CITYSCAPES DATASET

M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. [Bibtex]

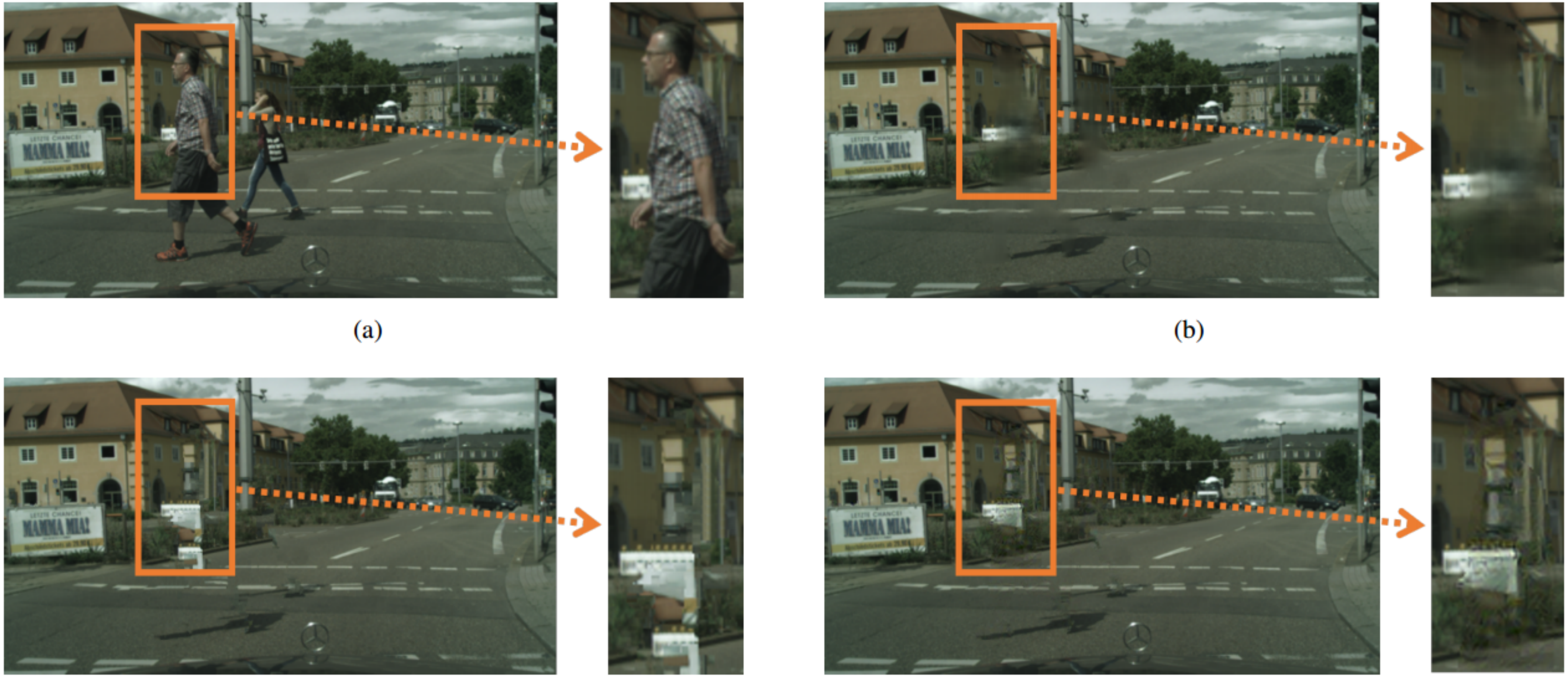
Type of annotations

1024 x 2048　　256 x 512

Generate Training Data – Pose Estimation



Generate Training Data - Inpainting

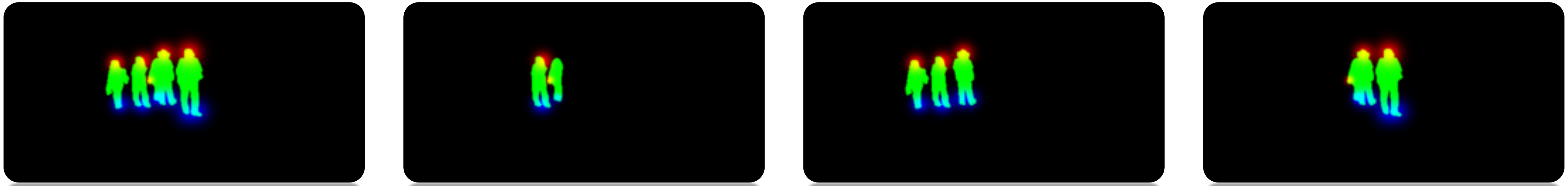


(a) (b) (c) (d)

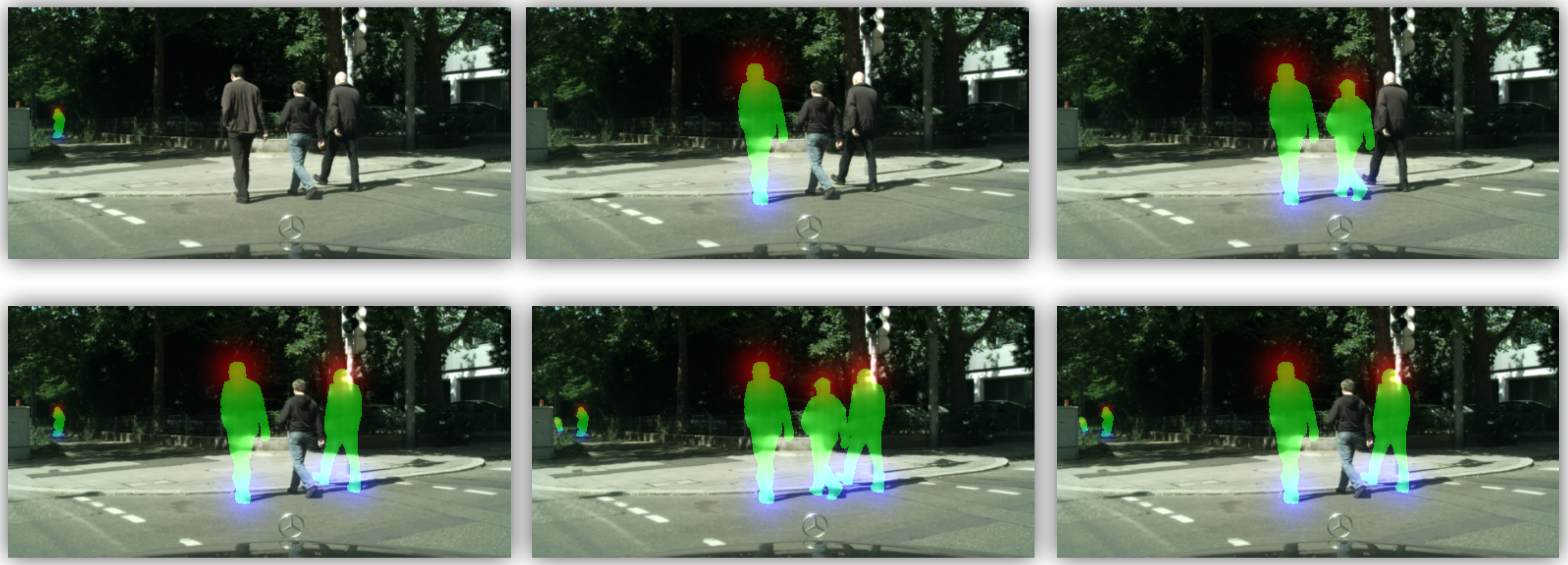
Generate Training Data – Input / Output

image

heatmap



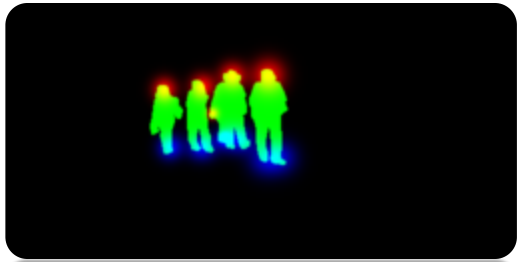
More Examples



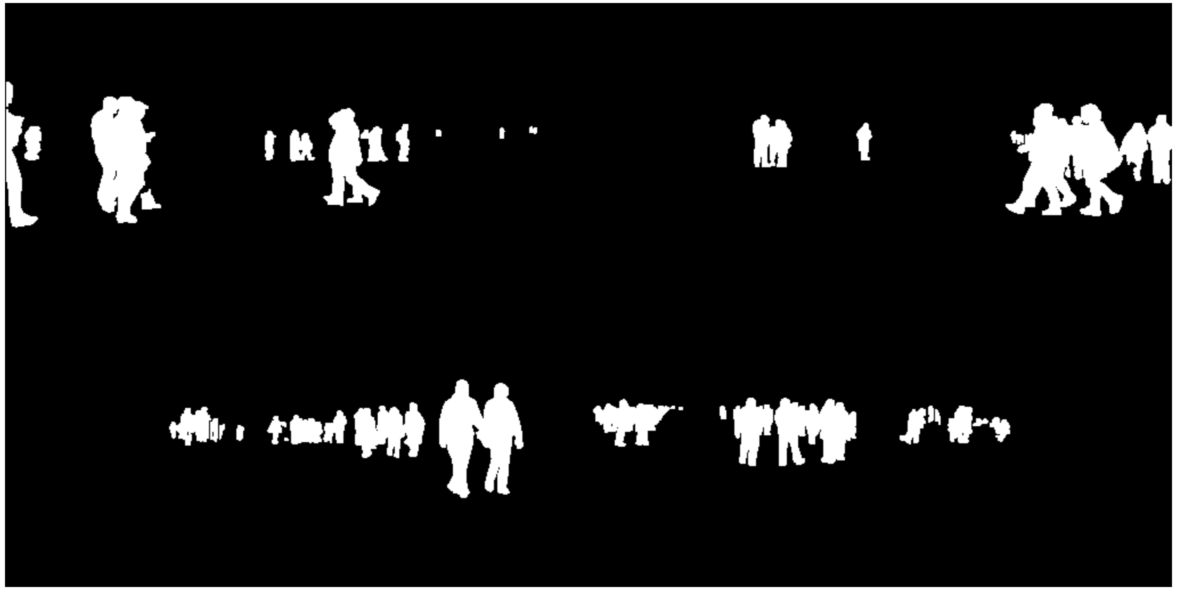
Pipeline

1. Generate training data
   - Collection
   - Pose estimation
   - Inpainting
   - Input / output
2. **Train the network**
   - **Adversarial learning**
3. Synthesize images
   - Pedestrian datasets
   - Synthesis

Task



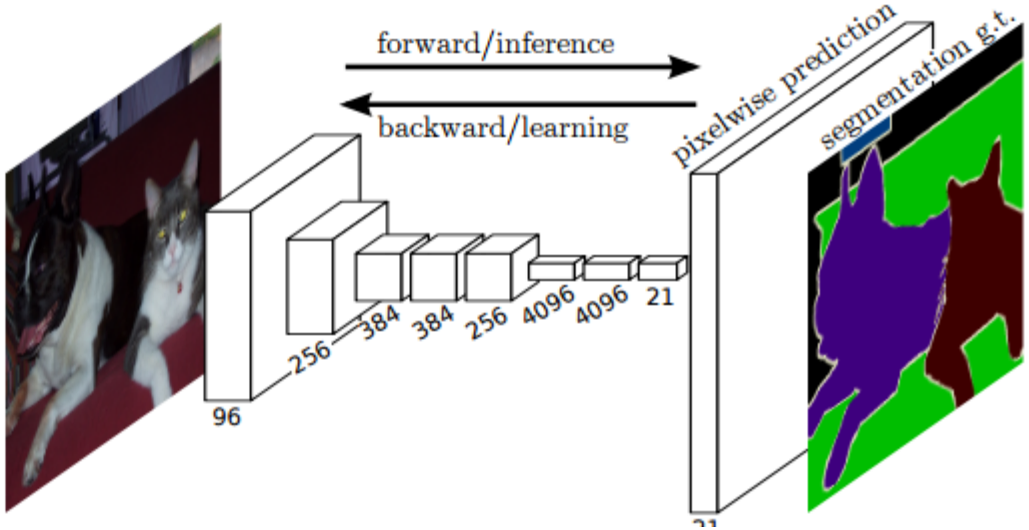
pedestrian pixel ratio > 5%



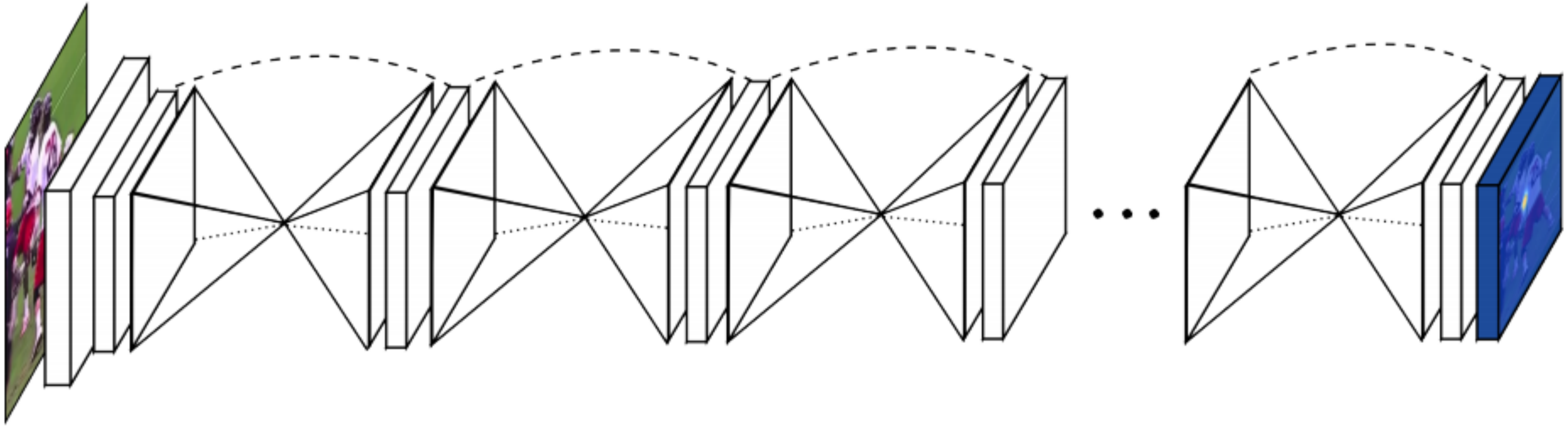
pedestrian pixel ratio <= 5%

Network

- FCN



- Stacked Hourglass



Adversarial Learning

- **GAN**
  - DCGAN
  - WGAN
  - WGAN improved
- **Using adversarial loss is popular**
  - Image-to-Image Translation with Conditional Adversarial Networks (CVPR2017)
  - Adversarial PoseNet: A Structure-aware Convolutional Network for Human Pose Estimation
  - SalGAN: Visual Saliency Prediction with Generative Adversarial Networks (CVPR2017 workshop)

And more…

- **Image Inpainting**
  - Context Encoders: Feature Learning by Inpainting (CVPR2016)
  - Generative face completion (CVPR2017)
  - Globally and Locally Consistent Image Completion (SIGGRAPH 2017)
- **Super-Resolution**
  - Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network
- **Object Detection**
  - A-Fast-RCNN: Hard Positive Generation via Adversary for Object Detection (CVPR2017)
- **Video Prediction**
  - Generating Videos with Scene Dynamics (NIPS2016)

Semantic Segmentation using Adversarial Networks

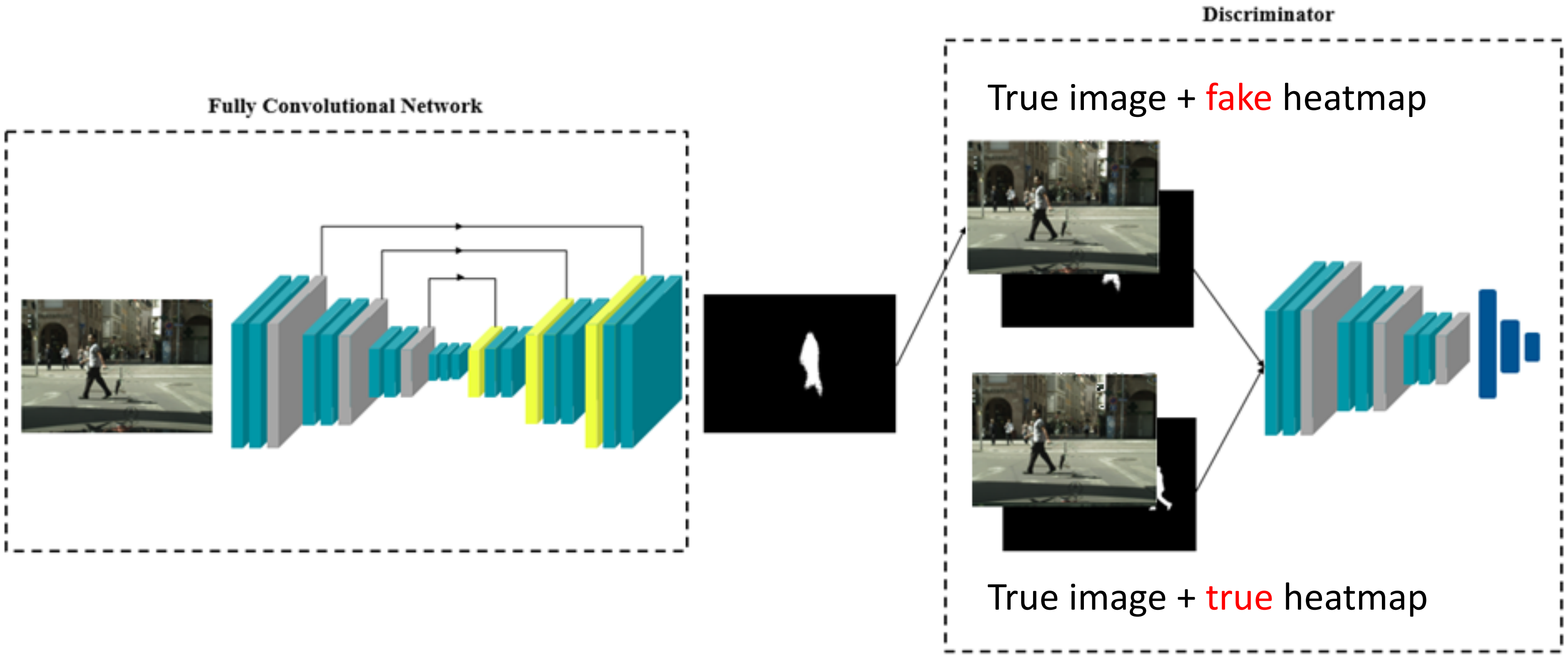
Reference:

Pauline Luc
Facebook AI Research
Paris, France
pauline.luc@fb.com

Camille Couprie
Facebook AI Research
Paris, France
coupriec@fb.com

Soumith Chintala
Facebook AI Research
New York, USA
soumith@fb.com

Ours:



Experiments - Validation

To evaluate the performance of these methods on predicting heatmaps, we use the recall rate as the metric. It is formulated as follows:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{area(l_i \cap h_i)}{area(l_i)}$$

where N is the number of images in validation set. $Area(l_i \cap h_i)$ is the overlap between ground truth heatmap l_i and predicted heatmap h_i.

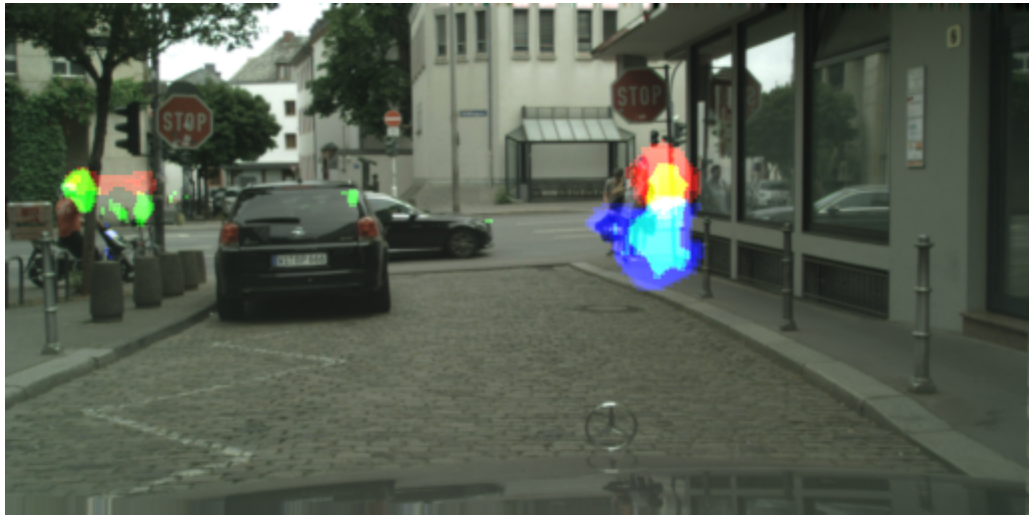
FCN	Hourglass	FCN+D
0.86	0.88	0.89

Experiments – Test

1. Sidewalks, safety islands, and bus stops are often assigned with high probabilities of pedestrian presence, even if the scene is void of pedestrians.
2. The timing is right: The *phantom* pedestrians is inclined to cross the street when there is no car.
3. People tend to form groups.
4. Depth and perspective are correct: The `sizes' of high-response areas in the heatmap are in accordance with the depth and vanishing point.

1. Sidewalks, safety islands, and bus stops are often assigned with high probabilities of pedestrian presence, even if the scene is void of pedestrians.

1. Sidewalks, safety islands, and bus stops are often assigned with high probabilities of pedestrian presence, even if the scene is void of pedestrians.

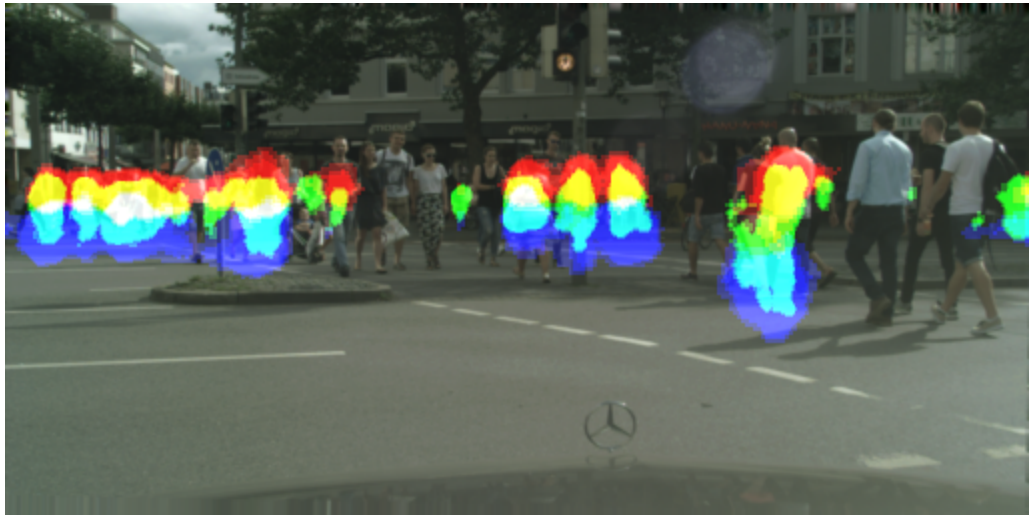


2. The timing is right: The `phantom' pedestrians is inclined to cross the street when there is no car.

2. The timing is right: The `phantom' pedestrians is inclined to cross the street when there is no car.

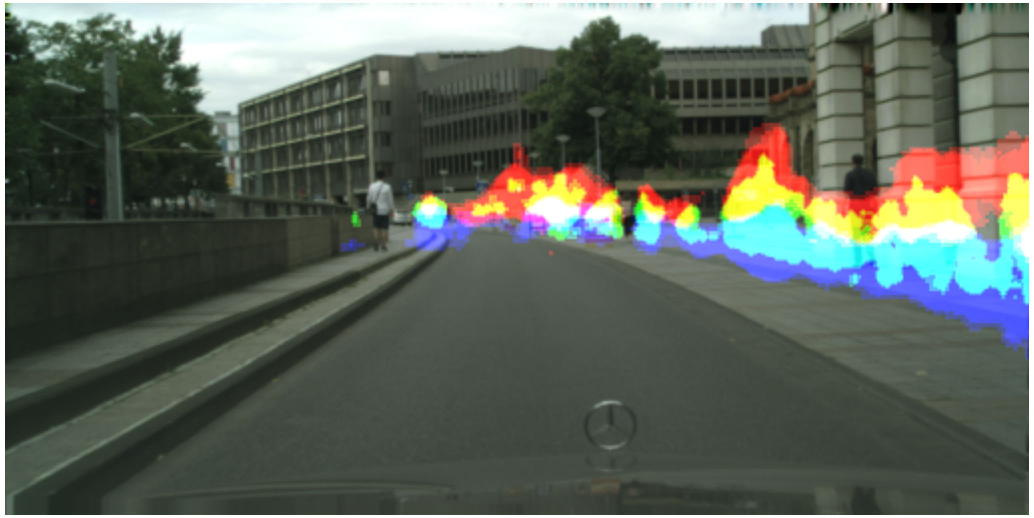
3. People tend to form groups.

3. People tend to form groups.

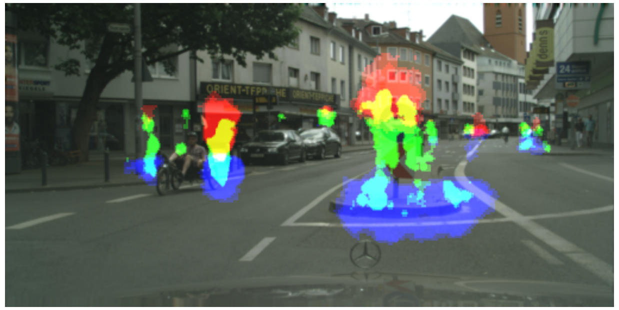
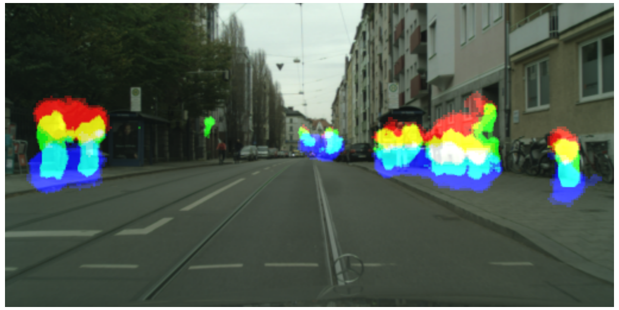
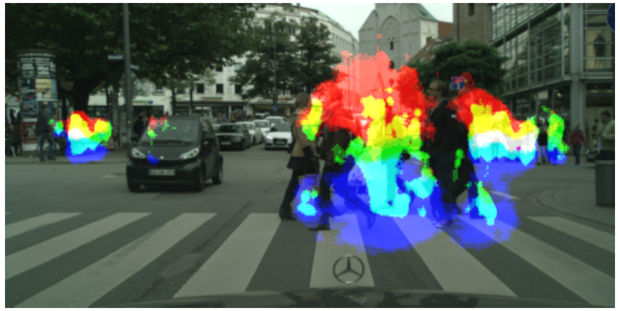


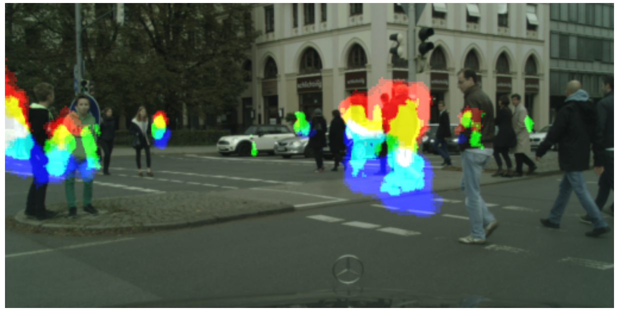
4. Depth and perspective are correct

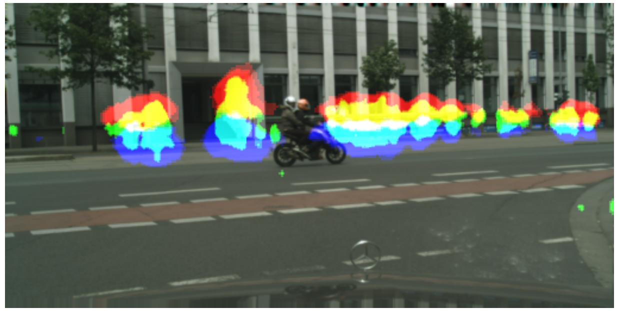
4. Depth and perspective are correct



More Results



Pipeline

1. Prepare training data
    - Collection
    - Pose estimation
    - Inpainting
    - Input / output
2. Train the network
    - Adversarial learning
3. **Synthesize images**
    - **Pedestrian datasets**
    - **Synthesis**

Pedestrian Datasets

- Images collected from *Cityscapes* and *PedCut*
- Categorized by height, width, aspect ratio…

Synthesis Pipeline - Basic

a. Input image

b. Predicted probability map (raw)

c. (b)with post processing

d. Synthesized according to (c)

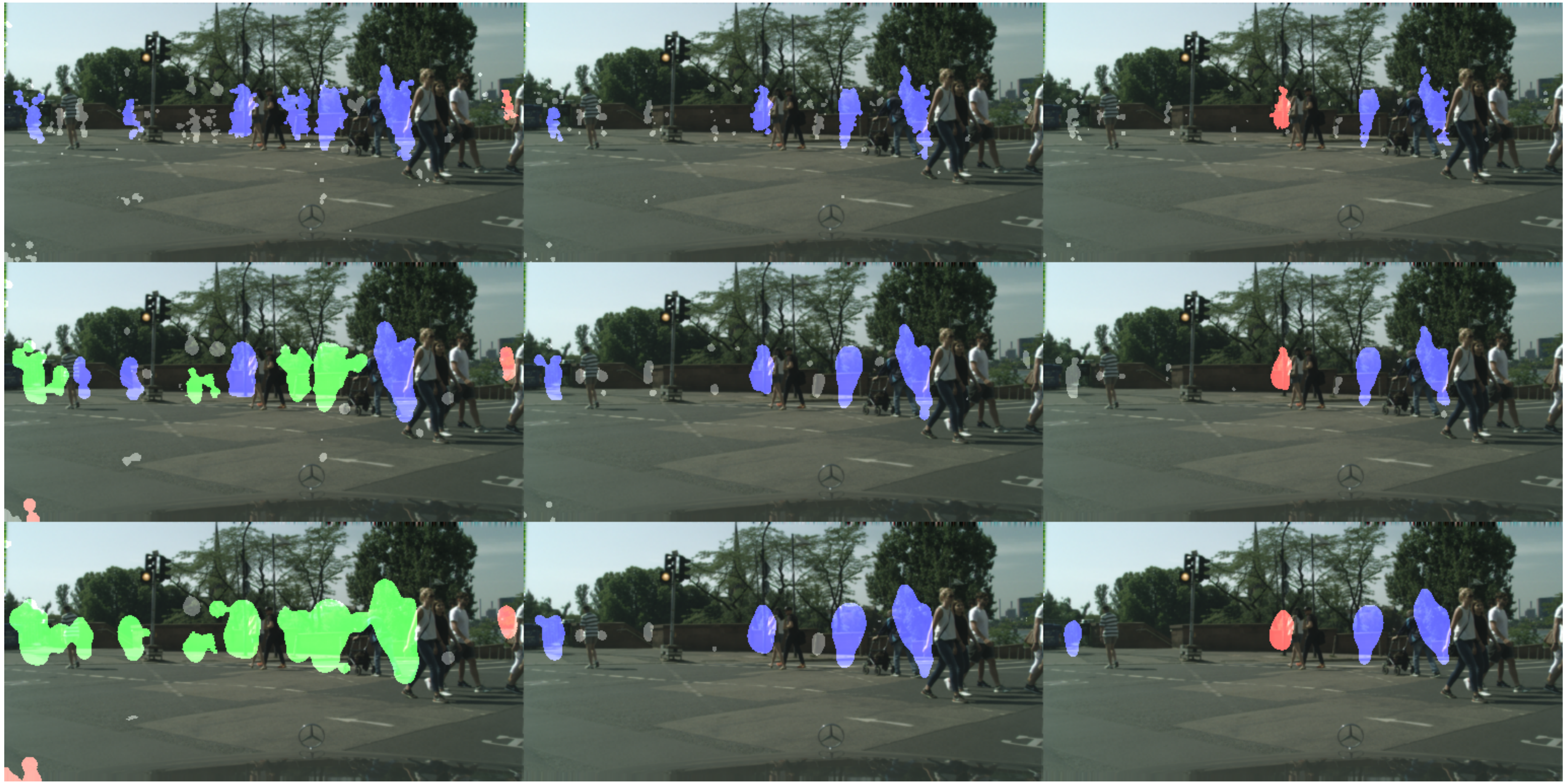
Hyperparameters

small **Threshold** big

small

smooth

big



Hyperparameters

small **Threshold** big

small

smooth

big

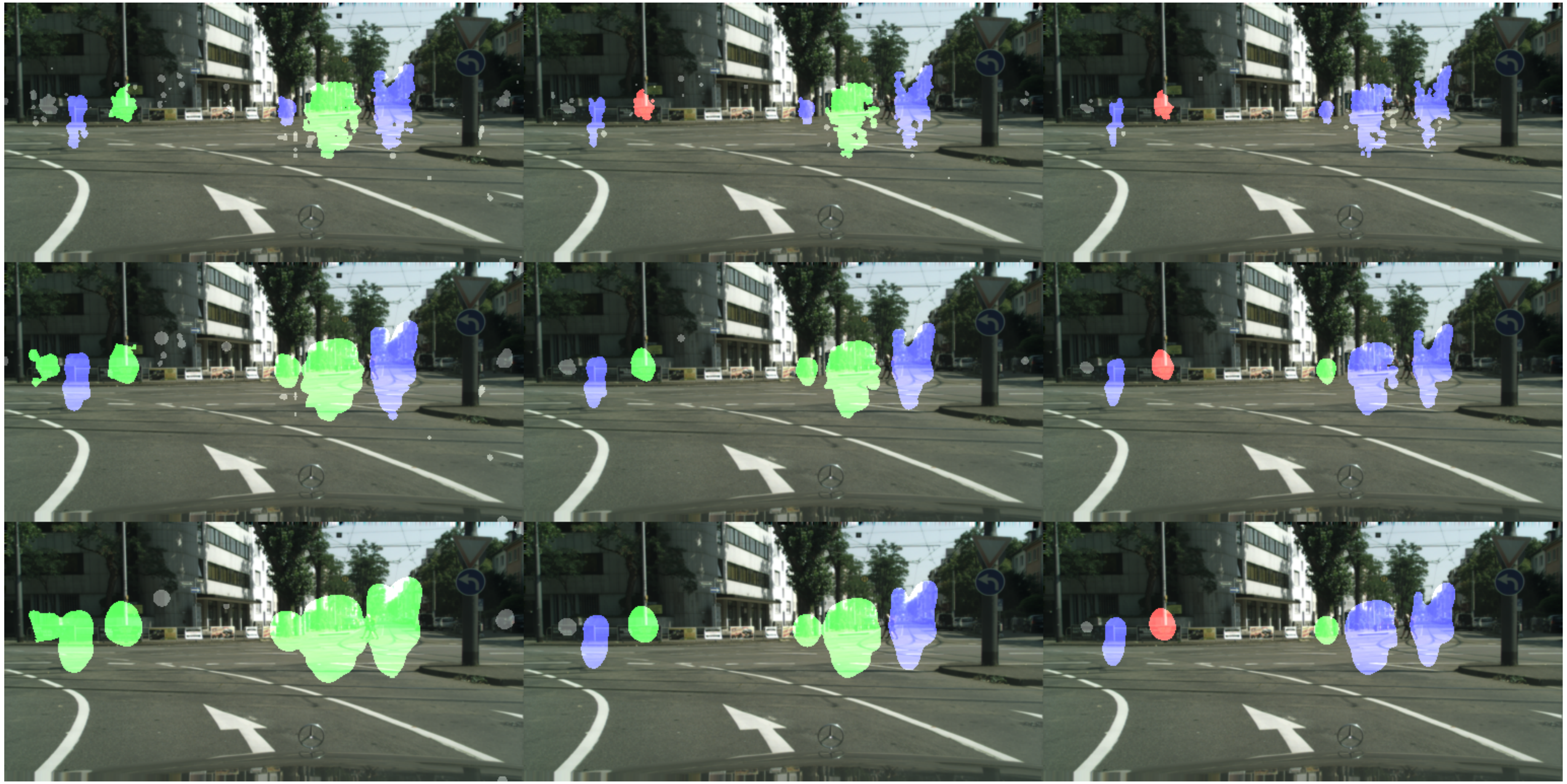
Hyperparameters

small **Threshold** big

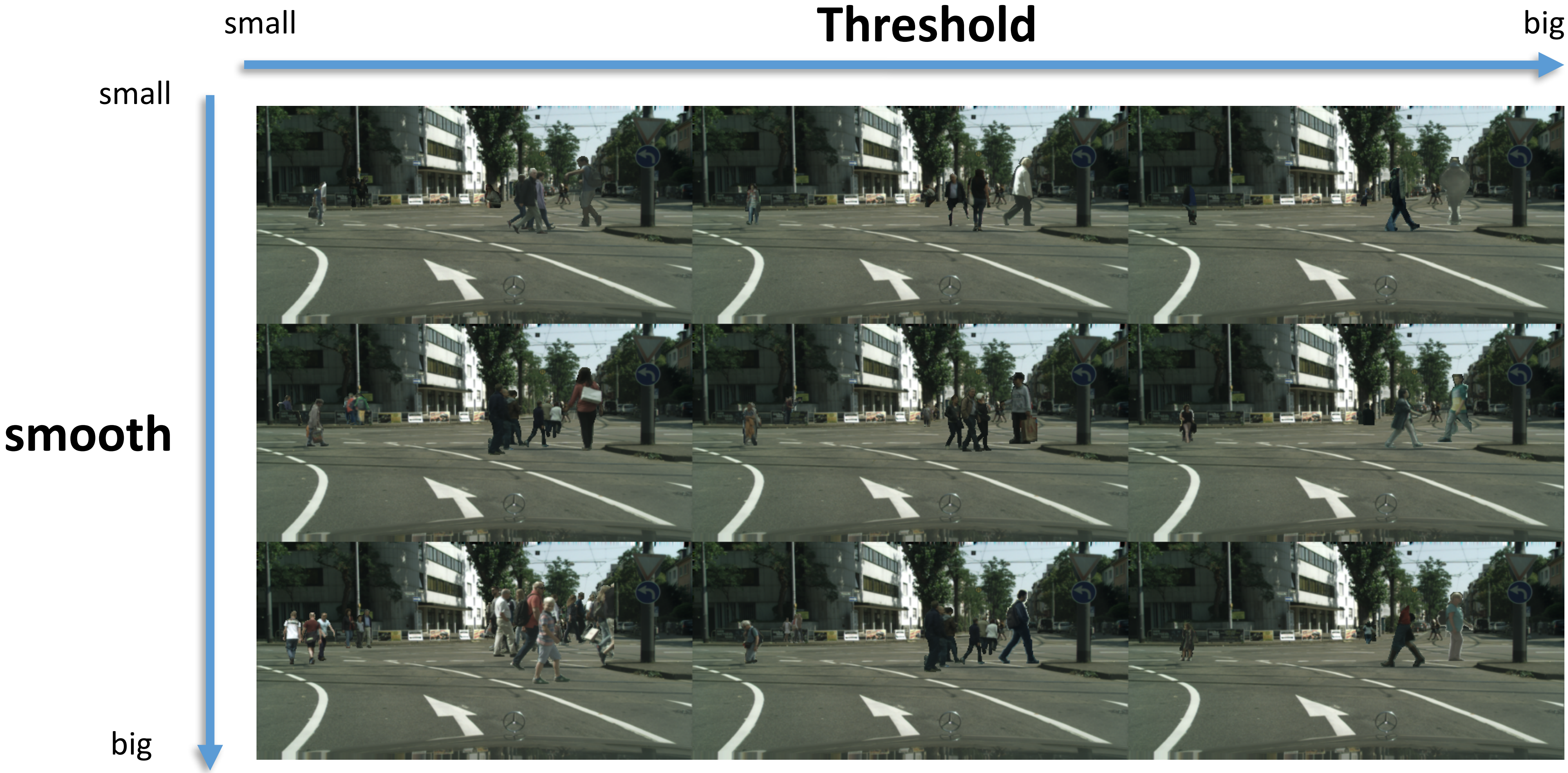
small

smooth

big

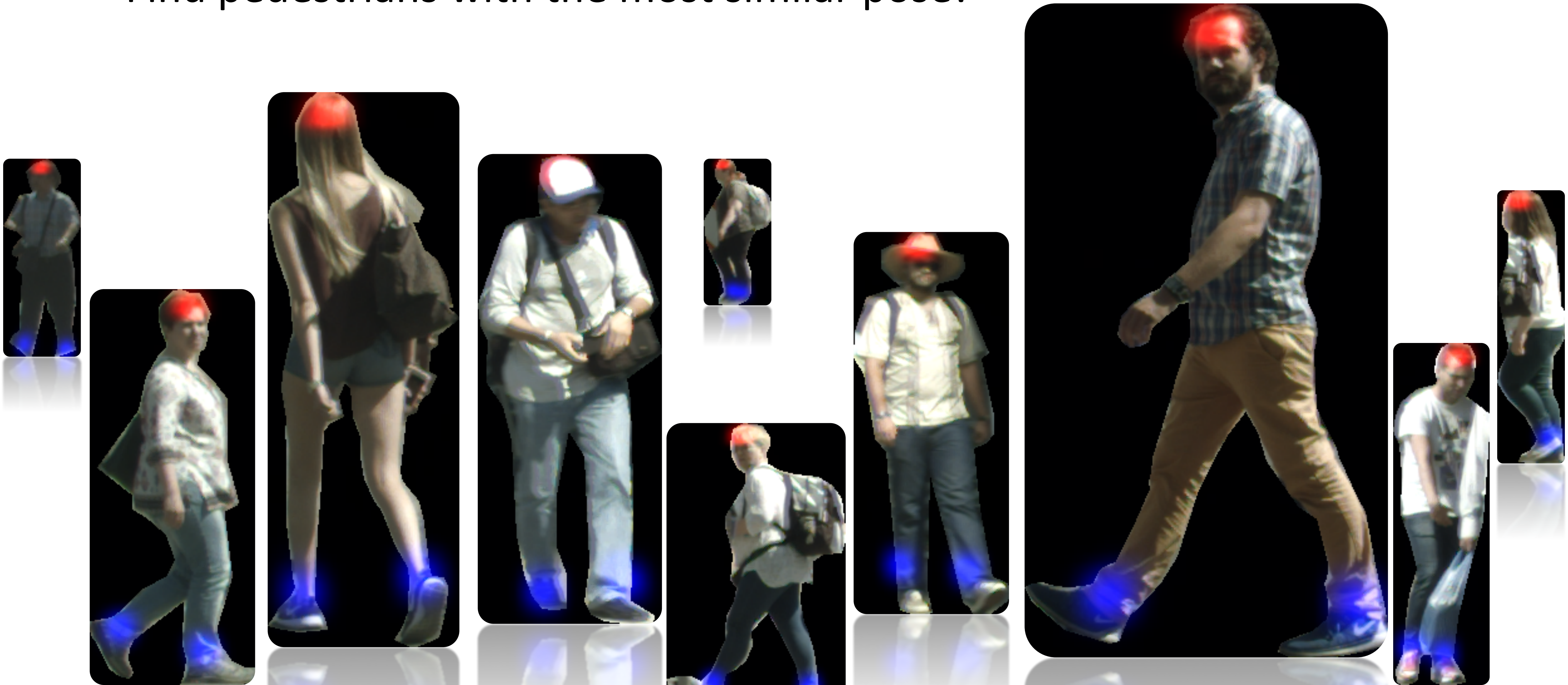


Hyperparameters



Synthesis Pipeline - Advanced

- Find pedestrians with the most similar pose!



More Results

Image | Probability map | Synthetic output

More Results

Image | Probability map | Synthetic output

More Results

Image | Probability map | Synthetic output

Thanks