Output of the Gaussian density discriminative network.

As expected the colors of the learned Gaussian classes in Fig. 3 match those predicted by the Network in Fig. 2. We see that our learning algorithm indeed identified several Gaussian classes which can be interpreted as "Left Head and Shoulders", "Right Head and Shoulders", "Right Flank" etc...

1 Detailed Explanation

This document illustrates the output of our Gaussian discriminative model and the learned Gaussian parameters.
The semantic segmentation of Fig. 1 corresponds to the output of the Background/Foreground segmentation Network. It is used in the "simple" occlusion model to compute

$$P^d(X_k = 0) = f_{\mathrm{b}}(\mathcal{F}^c_k; \theta_b) \ ,$$

and in the "full" model to compute

$$P^d(\vec{X}_k = 0) = f_{\mathrm{b}}(\mathcal{F}^c_k; \theta_b) \ .$$

Figure 1: Semantic segmentation $f_{\mathrm{b}}(\mathcal{F}^c_k; \theta_b)$.

The Network output of Fig. 2 represents the Gaussian class output by the network to compute $P^d(\vec{X}_k | \vec{X}_k \neq 0)$. More precisely, each pixel is colored proportionaly to,

$$f_{\mathrm{h}}(\mathcal{F}^c_k; \theta_{\mathrm{h}})_m (1 - f_{\mathrm{b}}(\mathcal{F}^c_k; \theta_b)) \ ,$$

where each Gaussian class m corresponds to a different RGB color. For convenience of representation, we only display 3 Gaussian classes out of 8 used in total.

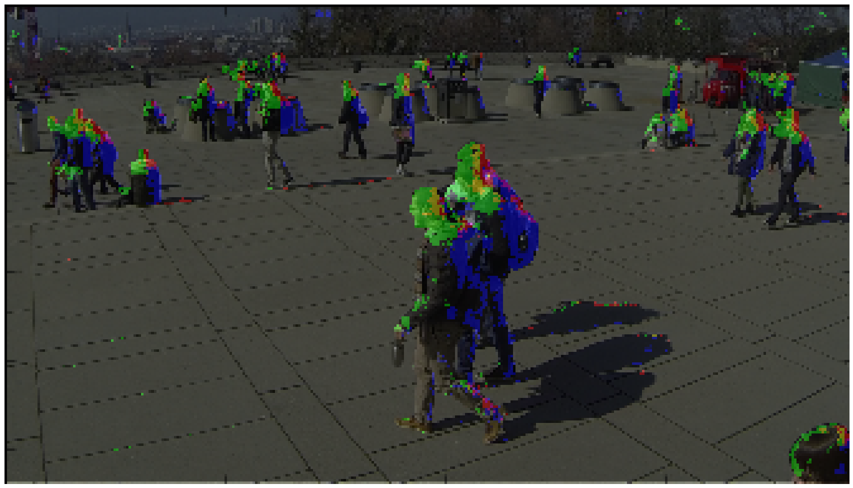

Figure 2: Gaussian Network output $f_{\mathrm{h}}(\mathcal{F}_k^c;\theta_{\mathrm{h}})_m(1 - f_{\mathrm{b}}(\mathcal{F}_k^c;\theta_b))$. We see that the pixels have been correctly identified as belonging to one of the three represented body-part, colored in *Red, Green and Blue*. We see that *Yellow* pixels appear, which correspond to pixels classified both as *Red* and *Green*.

Finally, we propose, in Fig. 3, to visualise the Gaussian parameters learned during training. To do so, we represent a projected bounding box centred in 0 of size $H \times W$. We then use a specific color to highlight the set of pixels which would vote for this bounding-box through each Gaussian element. For instance, the pixels k colored in *red*, where *red* corresponds to Gaussian class $m = 1$, are those such that,

$$\frac{\|x_k - \alpha_{1x}\|^2}{2(H\sigma_{1x})^2} + \frac{\|y_k - \alpha_{1y}\|^2}{2(W\sigma_{1y})^2} \leq 1 \; .$$

This representation has to be put in regards with the discriminative/generative correspondence. Indeed, let us assume that there is a single person in the image, with its single corresponding projected bounding-box in the camera plane. The discriminative model then matches the generative one if the network "colors" the pixels inside the bounding-box as in Fig 3.

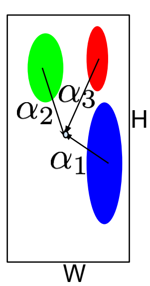

Figure 3: Representation of the learned three learned Gaussians with colors corresponding to the classes of Fig. 2.

2 Results

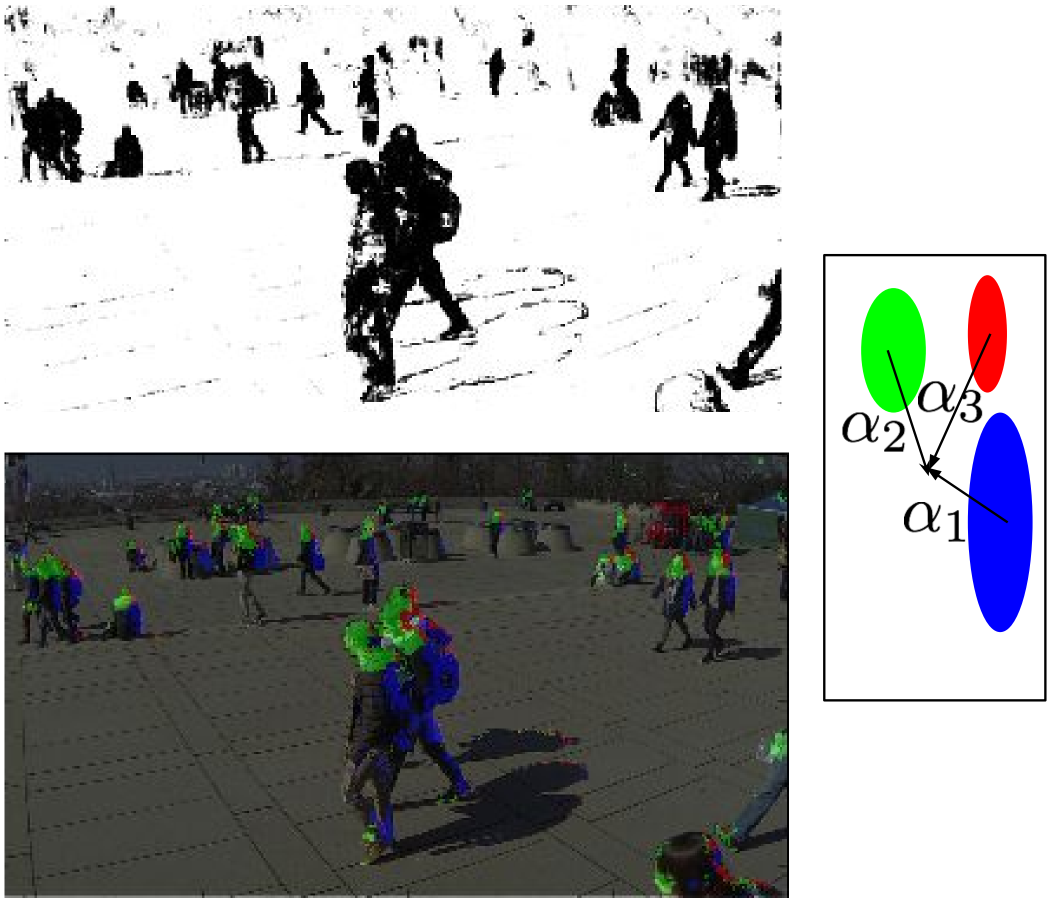

Figure 4: Camera 1. Gaussians 1,2 and 3 represented.

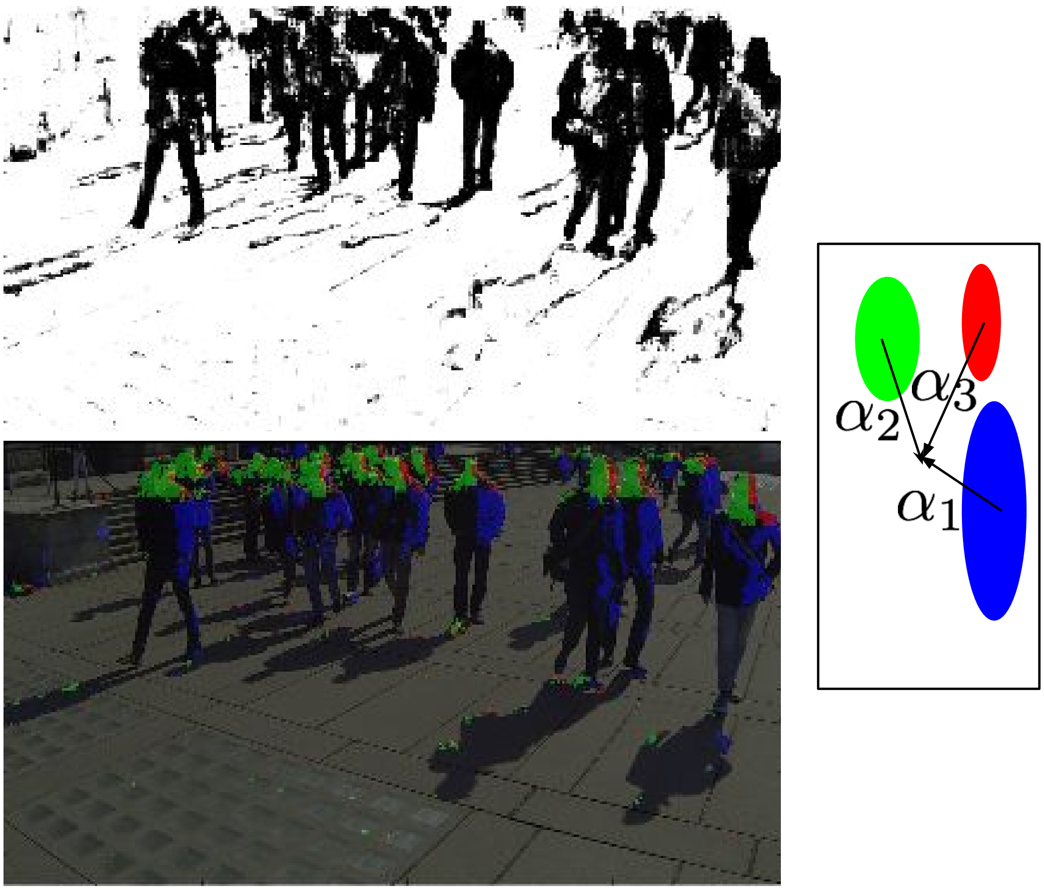

Figure 5: Camera 2. Gaussians 1,2 and 3 represented.

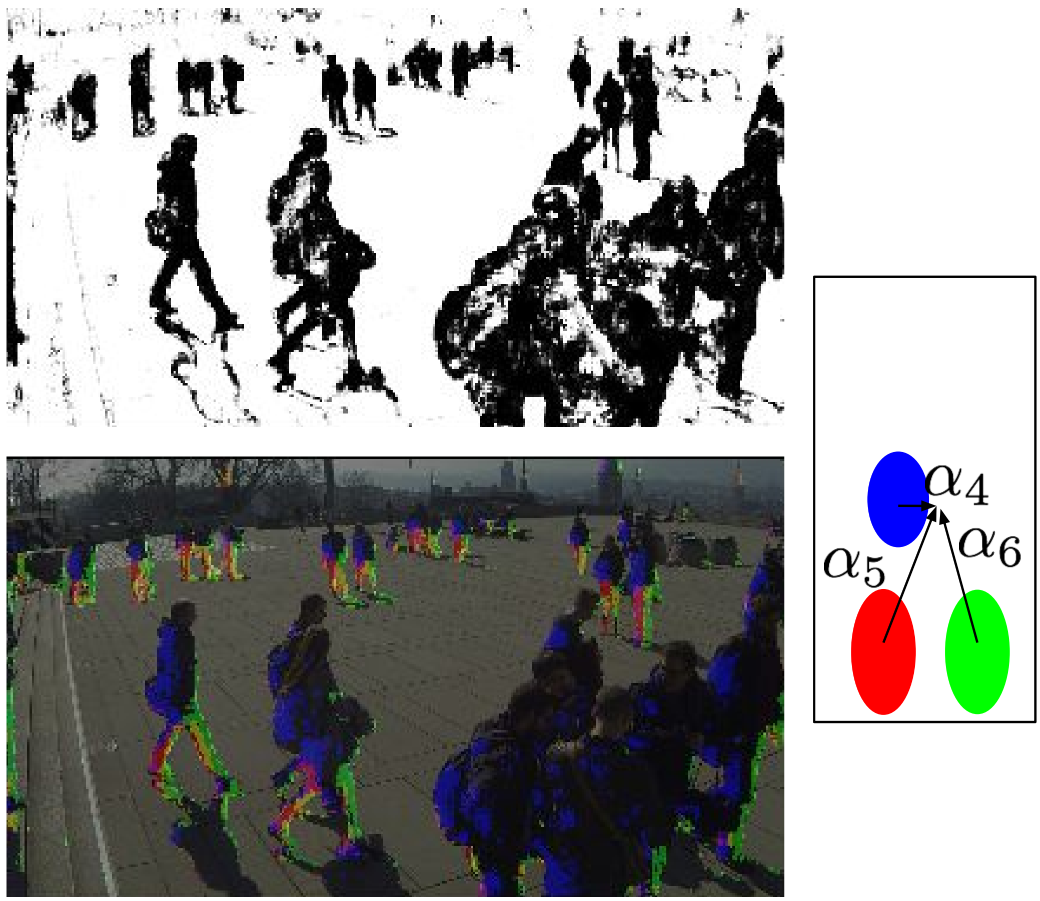

Figure 6: Camera 3. Gaussians 4,5 and 6 represented.

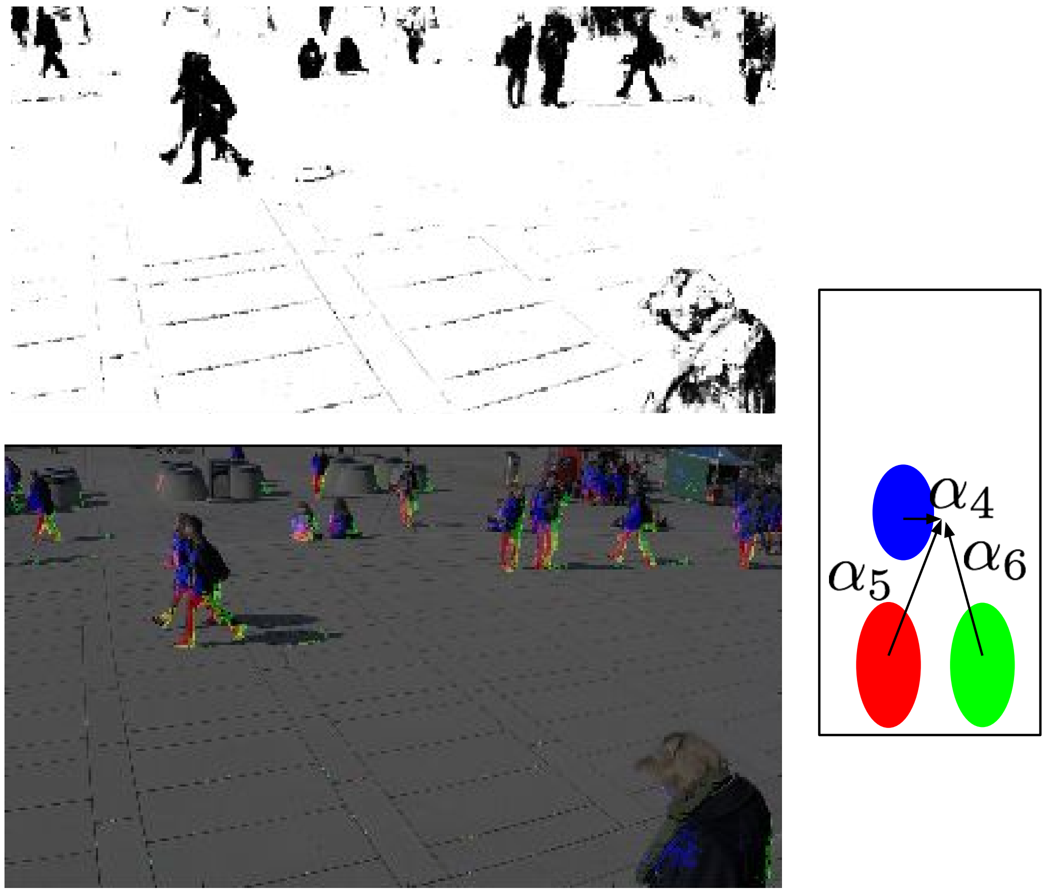

Figure 7: Camera 4. Gaussians 4,5 and 6 represented.