• **Goal:** recognizing pose from tiny images of people, down to 24px high
  • Tiny People are important in surveillance, autonomous driving, etc.
• **Challenge:** low resolution data is extremely ambiguous
• **Approach**
  • Models data uncertainty via probability distributions
  • A CNN that outputs a distribution over possible body joints from a small image of a person
  • A probabilistic variant of keypoint localization using dense heat maps
• **Evaluation**
  • Downsampling standard benchmarks (MPII Human Pose and MS-COCO)
  • Tiny People dataset of "real" low-resolution people

**MPII Human Pose dataset**
- Evaluation on the standard MPII Human Pose dataset\(^1\), where we reduced the resolution to approximate people seen at a distance

**Tiny People dataset**
- We introduce a new Tiny People dataset, which contains 200 real scene images capturing people at distance (the dataset is used for testing only)
- New Tiny People dataset of "real" low-resolution people

**Standard Formulation**
- The standard approach\(^2\) for landmark detection is to output one heatmap per joint
- Heatmaps are fitted to the ground truth data by a L2 per-pixel regression of a heuristic Gaussian-like kernel around the ground truth landmark location

\[
E_{\phi}(\mathbf{y}) = \sum_{i=1}^{N} \sum_{j=1}^{M} w_{ij} (y_{ij} - g_{ij}(\mathbf{y}; \phi))^{2}
\]

**Limitations:**
- Heatmaps only convey information about the location – they may appear to encode uncertainty, but they do not
- The accuracy is bound by the heatmap resolution

**Probabilistic Formulation**
- Our model emits a continuous Gaussian distribution for each keypoint \( \mathbf{u} \) by estimating Gaussian Mixture Model parameters over a coarse 16 × 16 feature map \( \Omega_d \) generated over the whole image

- Output: a continuous distribution over possible body joint configurations:

\[
p_{\phi}(\mathbf{u} | \mathbf{d}) = \sum_{\mathbf{v} \in \Omega_d} N(\mathbf{u} | \mathbf{v} + \Delta \mathbf{v}(\mathbf{d}), \Sigma(\mathbf{v}; \phi)) \cdot p_{\phi}(\mathbf{v} | \mathbf{d})
\]

- Training: minimum neg. log-likelihood of the ground truth keypoint locations

\[
E_{\phi}(\mathbf{y}) = -\frac{1}{N} \sum_{i=1}^{N} \log \prod_{j=1}^{M} p_{\phi}(y_{ij} | \mathbf{d})
\]

- The model allows for modelling data uncertainty (by increasing the variance of corresponding Gaussians), as well as for sub-pixel accuracy

### References

This research was supported by Continental