Bottom-Up and Top-Down Reasoning with Hierarchical Rectified Gaussians
Peiyun Hu, Deva Ramanan

Core ideas
- Human vision suggests that feedforward processing suffices for vision-at-a-glance tasks (scene recognition), but feedback is needed for vision-with-scrutiny (fine-grained spatial understanding) [Hochstein and M. Ahissar, Neuron 02].
- We incorporate feedback into CNNs by unrolling inference on a hierarchical probabilistic model.
- Our model adds top-down feedback “for-free”, without increasing the number of parameters in standard networks.
- Near state-of-the-art results for facial and human landmark localization.

Approach
- We make use of Rectified Gaussian models [Socci et al, NIPS 98], a marriage of Boltzmann machines and Gaussians.
- Socci et al demonstrate that MAP inference on such models corresponds to a quadratic program (QP).
- We show that coordinate QP updates can be unrolled into a rectified recurrent net that naturally incorporates bottom-up and top-down processing.
- We train our inference engine with backprop using standard CNN packages.

Heatmap predictions improve by adding lower-level features (through skip connections) and top-down feedback (that fixes poor localization of knee and left/right ankle ambiguity).

Top-down feedback is particularly helpful when localizing occluded landmarks.