Paper link:

DeepGender: Occlusion and Low Resolution Robust Facial Gender Classification via Progressively Trained Convolutional Neural Network with Attention

Felix Juefei Xu*, Eshan Verma*, Parag Goel, Anisha Cherodian and Marios Savvides

felixu@cmu.edu, everma@cmu.edu, msavvid@ri.cmu.edu

IEEE 2016 Conference on Computer Vision and Pattern Recognition

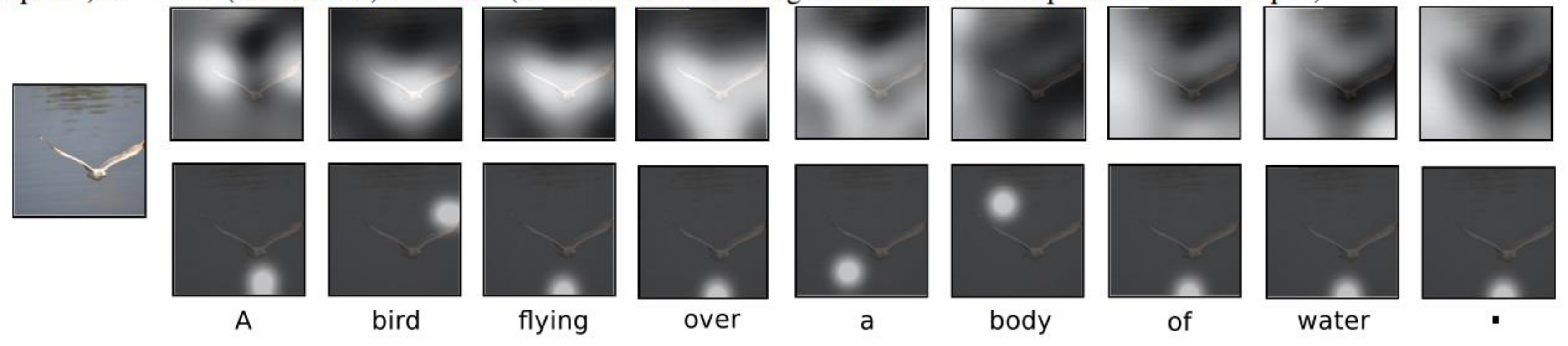
CVPR 2016

Motivation

Attention based model [1] can automatically fix its gaze on the salient objects (regions) in the image while generating the image caption word by word.

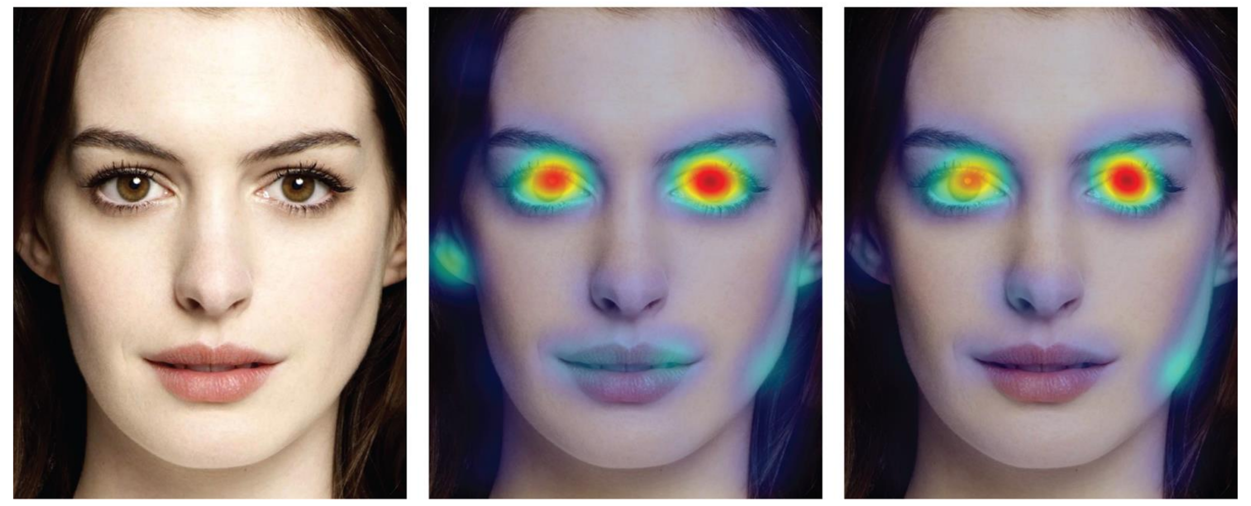
Q: Can we control / enforce the attention shift in CNN?:

Figure 2. Attention over time. As the model generates each word, its attention changes to reflect the relevant parts of the image. "soft" (top row) vs "hard" (bottom row) attention. (Note that both models generated the same captions in this example.)



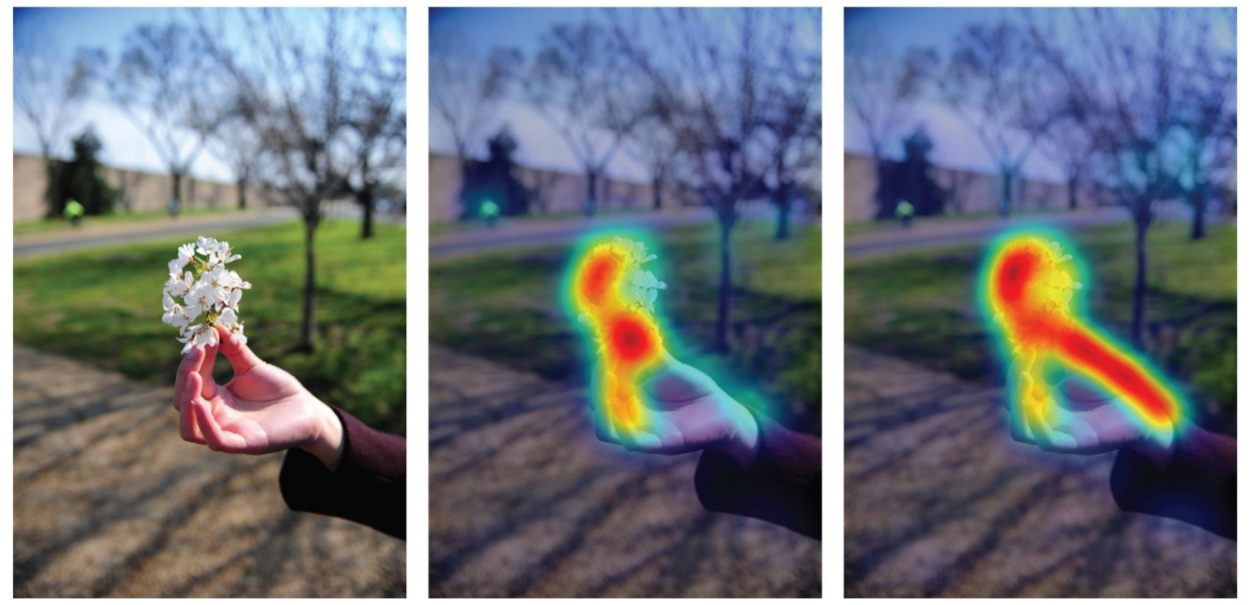
The periocular region has been shown to provide the most important cues for determining gender information. The periocular region is also the most salient region on human faces.

Input Image | Image Signature - LAB | Image Signature - RGB



Q: How can we let the CNN shift its attention towards the periocular region?:

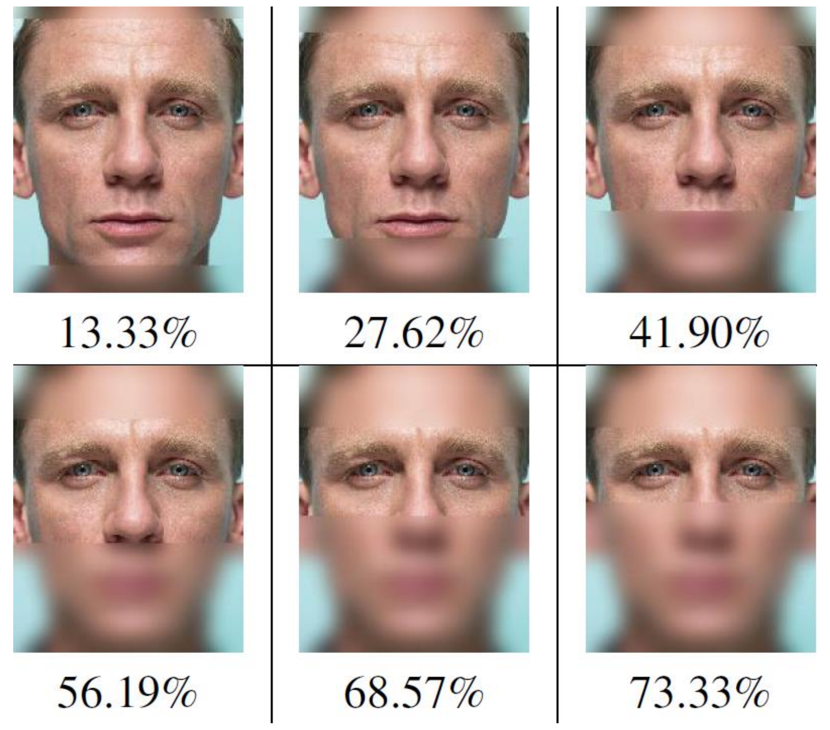
Input Image | Image Signature - LAB | Image Signature - RGB

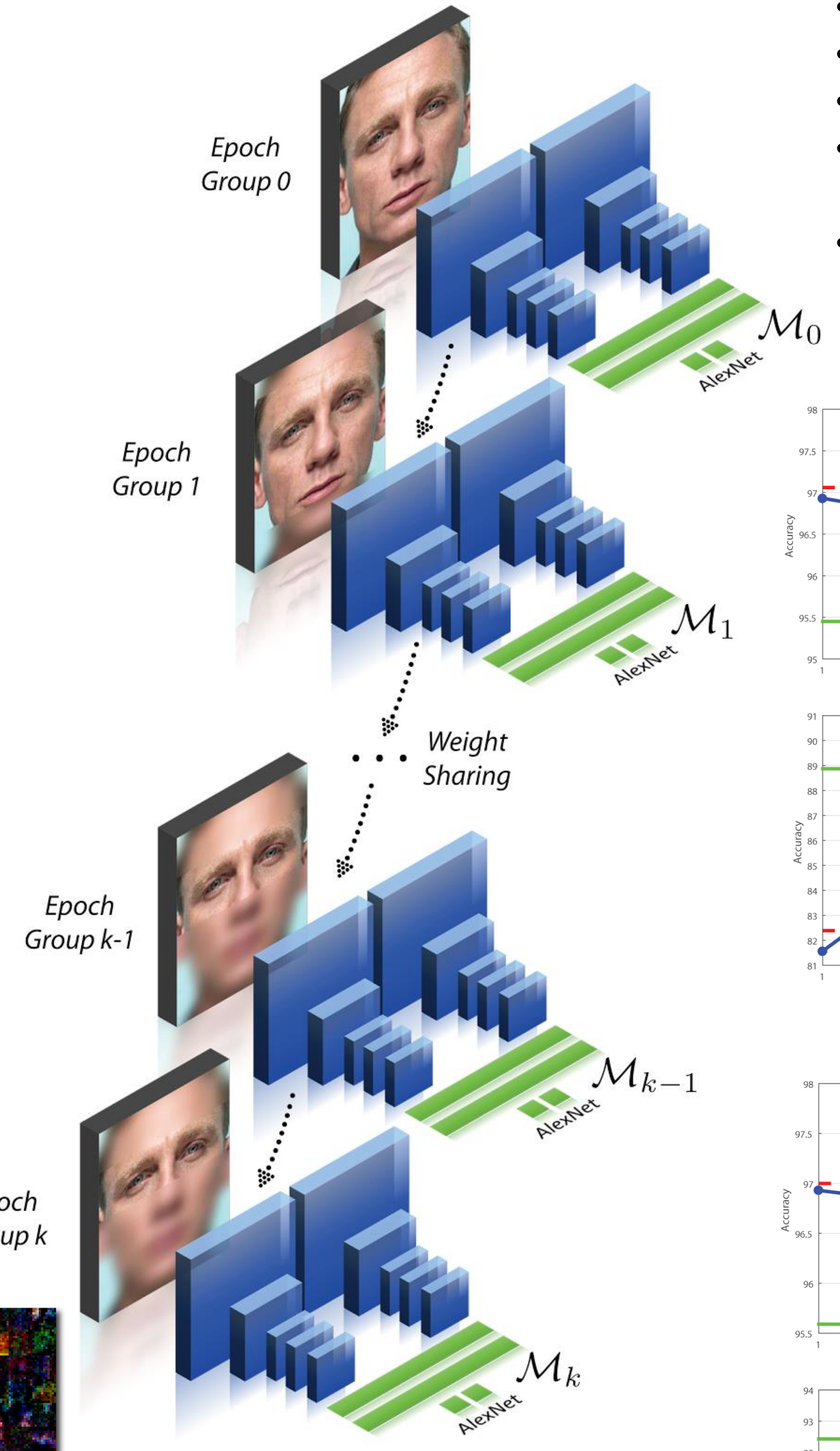


[1] K. Xu, et al, "Show, attend and tell - neural image caption generation with visual attention", ICML 2015.

Approach

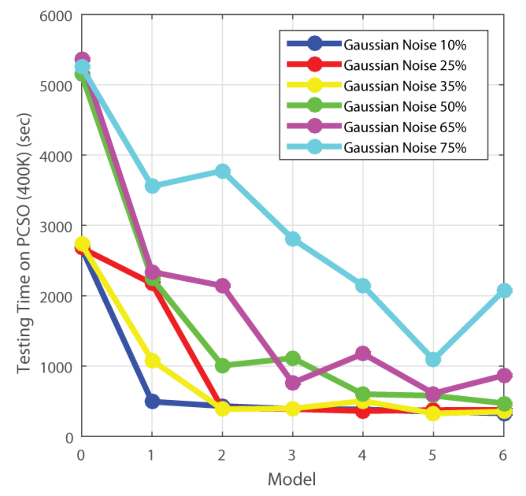
Enforcing attention in the training images: We heuristically choose 7 blur levels. Gaussian blur kernel with $\sigma = 7$.

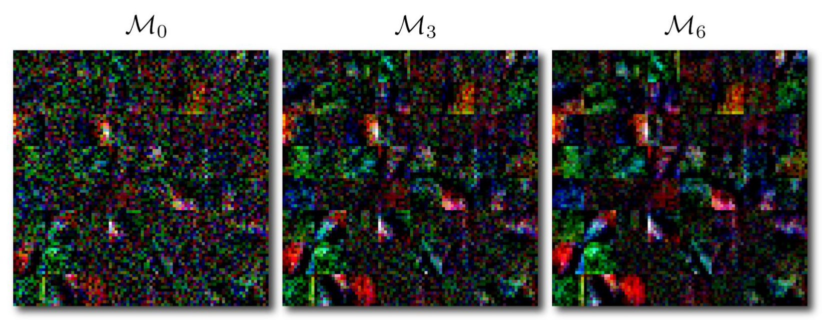




Progressively trained CNN with attention:

- Training starts with the first epoch group (Epoch Group 0, images with no blur), and the first CNN model M_0 is obtained and frozen after convergence.
- Then we input the next epoch group for tuning the M_0 and produce M_1.
- AlexNet with 2-way softmax.
- Each M_j (j=0,…,k) is trained with 1000 epochs with batchsize of 128.
- progressively training the CNN using blurred images serves as an implicit low-rank regularizer.





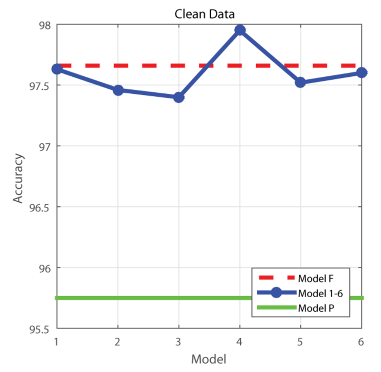
Carnegie Mellon University
CyLab Biometrics Center

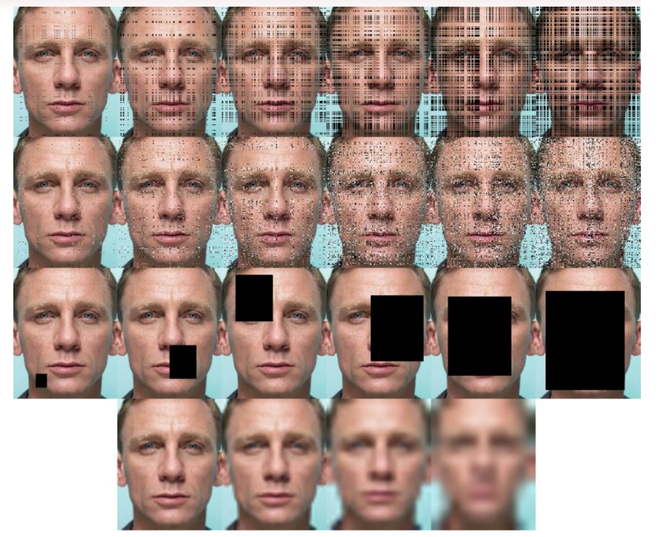
Experiments

Pre-processing on 400K PCSO testing images:

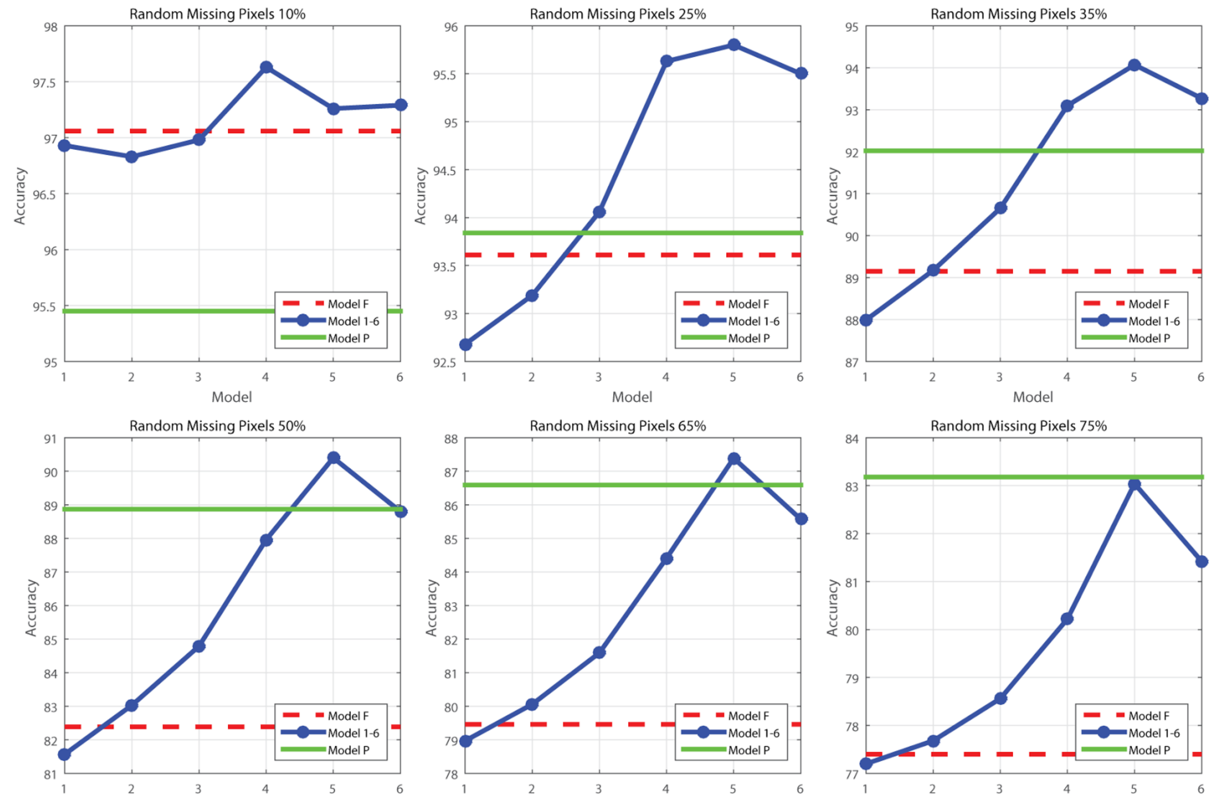
- Row 1: random missing pixel occlusions
- Row 2: random additive Gaussian noise occlusions
- Row 3: random contiguous occlusions
- Percentage of degradation for Row 1-3: 10%, 25%, 35%, 50%, 65%, 75%.
- Row 4: various zooming factors (2x, 4x, 8x, 16x) for low-resolution degradations

Baseline:

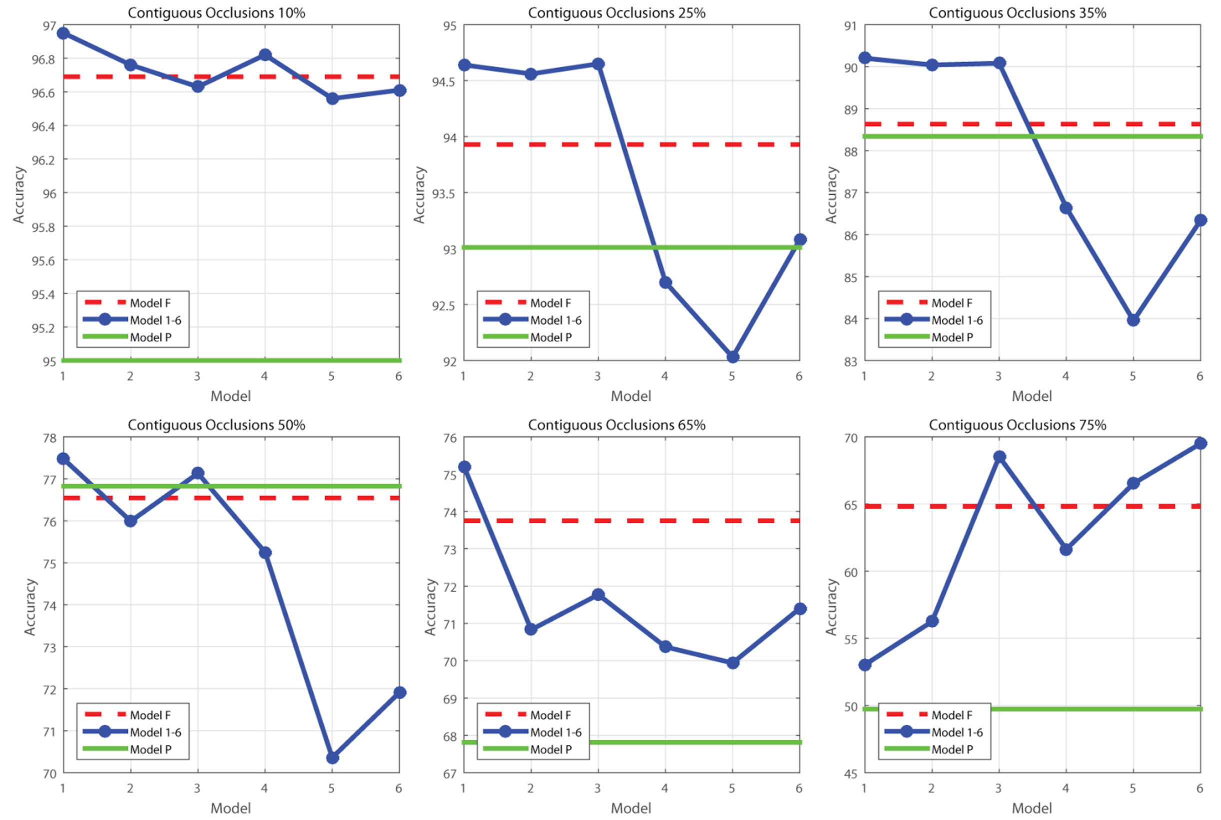




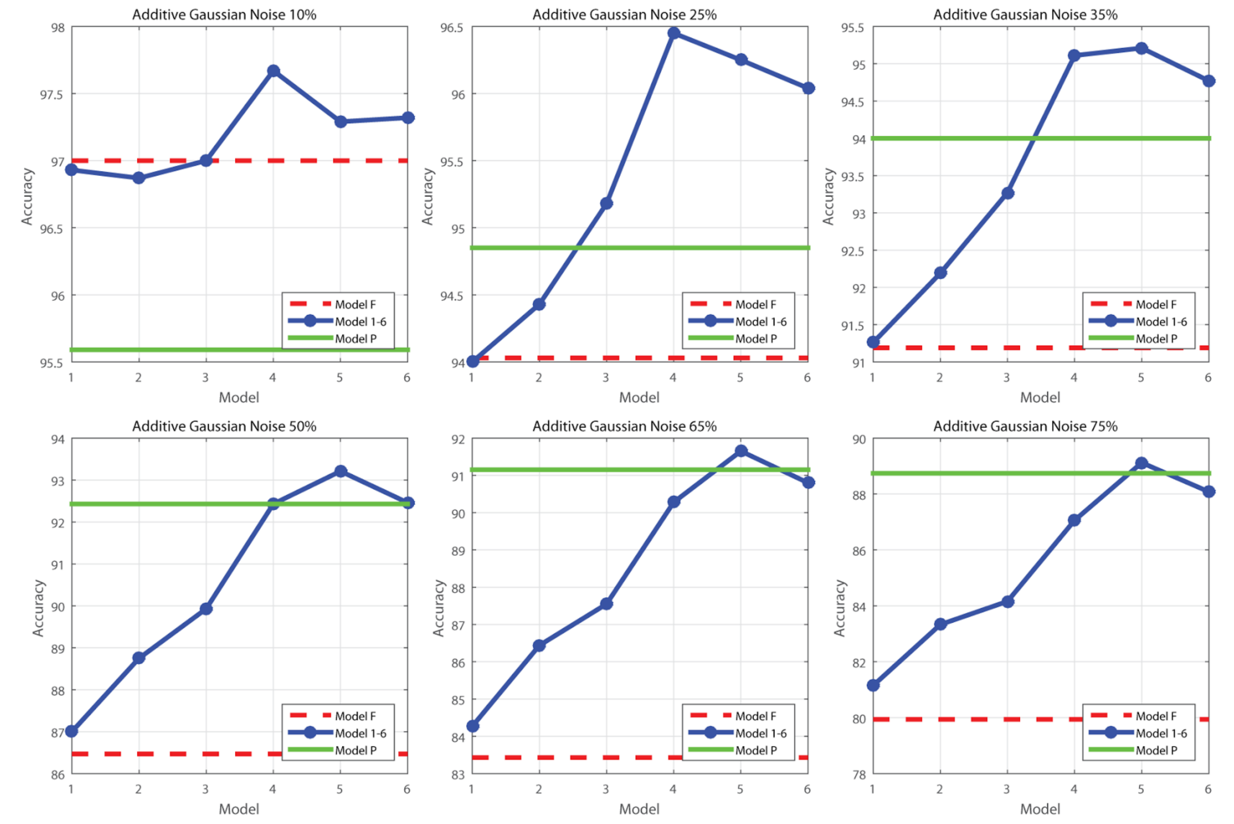
Random missing pixels occlusions:



Random additive Gaussian noise occlusions:



Random contiguous occlusions:



Low-resolution degradations: