Unsupervised Representation learning by predicting image rotations

Summary

1. Train Convnet to predict the 2d rotation that is applied to the image it gets as input.

2. Demonstrate dramatic perf improvement compared to prior UL methods.

3. 2 days on a single Titan X GPU.

Predicting Rotations

1. 0, 90, 180, 270 degrees.

2. The ConvNet has to predict the transformation without having access to the original untransformed image.

3. Intuitive argument: it is not possible to perform this task.

   without knowing how to:
   - recognize classes of the objects in the image
   - ~~the obje~~ detect the semantic parts of the object

4. More concretely, the model needs to:
   - localize salient objs in the imgs
   - recognize their orientation and type
   - relate the obj's orientation with the dominant orientation that each type of objects tend to be depicted within the available images.

5. Argument wrt 4 prior methods: image rotations do not create low-level visual artifact that allow the network to cheat in the task