Learning by Imagination: Parsing Animals

- A limitation of most of AI vision is that it relies heavily on models trained on annotated datasets. It is biologically implausible that annotations are available to humans (particularly infants). Instead humans develop models of vision over a period of a few years with stereotypical stages of development (e.g., binocular stereo develops between the ages of 4 and 14 weeks). Even for computer vision, some annotations are very hard or impossible to obtain.
- An alternative approach is learning by immagination. Suppose we have an internal three-dimensional models of an object (e.g., a toy panda, horse, or tiger) which can manipulate in our head and simulate with different poses, lighting, and textures. This can be mimiced using computer graphics rendering tools. For these ”internal models” we will know the three-dimensional positions of the joints/parts. Hence a large amount of synthetic data can be simulated and used to train a deep network models (or any other model).
- This model can then be transferred to work on real data. This is non-trivial since real world data (real tigers and horses) have different texture properties. But this can be done, particularly if video sequences are available, by exploiting temporal consistency (see handout).
- There is no evidence that humans learn models of objects in this way. But there is plenty of evidence that humans have internal mental representations (Kosslyn, Shepard, etc). and plausible arguments that ”we have a physics simulator in our heads” (Tenenbaum).