Unsupervised Deep Networks

- Deep Networks require an enormous amount of detailed supervision provided by large annotated datasets. This is a problem for AI, because annotating datasets can be extremely time-consuming and sometimes impossible. Also it is inconceivable that humans learn by detailed supervision.
- We describe two strategies: (i) learning using the loss function, (ii) bootstrap learning, where learning some visual skills enable the unsupervised learning of other skills.
- The lecture should also have mentioned some alternative approaches such as learning by jigsaw puzzles (where the deep network is trained to discriminate between a shuffled image and a normal image). The handout by Chen Wei gives a entry point to this literature.
- The lecture should also have mentioned few-shot learning. This shows that features vectors learnt for some Deep Network tasks can be used with few examples (i.e. limited annotation) to learn other tasks. The handout by Qiao et al. gives an entry point to the literature.

Unsupervised Deep Networks: Learn from the Loss Function

- One strategy for unsupervised learning is to use a loss function which imposes prior assumptions about the data.
- For example, if the goal is to train a Deep Network to do image smoothing and edge detection, then we can assume that the images are weakly smooth. This is modeled by setting the loss function to be the energy function of a Markov Random Field model (see earlier lectures). This means that the Deep Network does not need to know the answer (i.e. it doesn't need to be fully supervised) but it does have to know the form of the solution (i.e. that is weakly smooth).
- Another perspective is that the Deep Network learns to do inference on a Markov Random Field model. This is a much faster algorithm (almost instantaneous) than using mean field theory or Gibbs sampling.
- The handout by Smirnakis and Yuille is (to the best of my knowledge) the first work of this type. It requires no supervision but does need training data of images (hard to get when we did the work). The handout by Zhe Ren describes very recent work that uses this strategy for learning to estimate motion (optical flow) with a prior that neighboring pixels have very similar motions (see early lecture on motion.
- Note: this type of learning assumes that the loss function (or MRF model) is known (so there is some supervision). The studies on estimating motion have results which are almost as good as supervised methods.

Unsupervised Deep Networks: Bootstrap Learning

- A complimentary strategy is to learn deep networks for one visual task and using their results to train deep networks for other visual tasks. This is a bootstrapping strategy.
- Here is an example. An agent (a car) moves within a domain (city road scenes) which consists of a static background (roads and buildings) with a few moving objects (humans and other cars).
- One Deep Network can learn motion estimation (using the methods in the first slide). A second deep network can learn the depth of the static background by exploiting the fact that agent (car) is moving rigidly relative to the background (this should be discussed in the course, it exploits factorization models (e.g., Kontsevitch et al, Tomasi and Kanade). These depth estimates (from the second deep network) will be wrong for objects that are moving relative to the background, but this can be exploited to find the moving objects (by finding inconsistencies between the motion estimated by the first deep network and estimates which follow from the second deep network).
- This can be extended a third deep network that uses the depth from the second deep network as supervision. Note: this only works for objects which have been seen statically (i.e. in the background).
- The strategy, and its relations to classical ways to combine different visual tasks (cues), is discussed further in the handout Bootstrapping Deep Networks. Some more details are given in the handout on Joint Learning of Geometry and Motion.