TTIC 31230 Fundamentals of Deep Learning

Problems for CTC.

Problem 1. This problem is on a CTC-like algorithm for image labeling.

Suppose that the training data consists of pairs (I, S) where I is an image and S is a set of object types occuring in the image. For example S might be $\{\text{Person}, \text{Dog}, \text{Car}\}$. To be concrete we can take $\mathcal{C}$ to be the set of image labels used in CIFAR 100 and take S to be a subset of $\mathcal{C}$ containing no more than five labels ($|S| \leq 5$). We want to do SGD on a model defining $P_\Phi(S \mid I)$.

We will use a latent variable $z[X, Y]$ such that for pixel coordinates (x, y) we have $z[x, y] \in \mathcal{C} \cup \{\perp\}$. For a given $z[X, Y]$ define $S(z[X, Y])$ to be the set of classes appearing in $z[X, Y]$, i.e., $S(z[X, Y]) = \{c \; \exists \; x, y \; z(x, y) = c\}$. Here the "semantic segmentation" $Z[X, Y]$ is analogous to the phoneme sequence $z[T]$ in CTC. Unlike the CTC model, the label S is a set rather than a sequence.

We assume a CNN (with convolutions of stride 1 to preserve spatial dimensions) followed by a softmax at each pixel to get a probability $P_\Phi(z[x, y] = c)$ for each pixel location (x, y) and each $c \in \mathcal{C} \cup \{\perp\}$ and where each pixel location has an independent probability distribution over classes. To simplify notation we can reshape the pixel locations into a linear sequence and replace $z[X, Y]$ by $z[T]$ with $T = X \times Y$so we have $z[0], z[1], \ldots, z[T - 1]$.

Define

$$S_t = \{c \in \mathcal{C} \; \exists t' \leq t \; z[t'] = c\}$$

For $U \subseteq S$ define

$$F[U, t] = P(S_t = U)$$

Note that for $|S| \leq 5$ there are at most 32 possible values of U. Give dynamic programming equations defining $F[U, 0]$ and defining $F[U, t + 1]$ in term of $F[U', t]$ for various U'.