Learning a Deep Compact Image Representation for Visual Tracking

mingzailao

2016-9-11

Outline

1. Training
2. Experiments

Offline Training with Auxiliary Data

Dataset

Tiny Images dataset : 80 million tiny images each of size 32×32

- randomly sample 1 million images for offline training

Offline Training with Auxiliary Data

Learning Generic Image Features with a Stacked Denoising Autoencoder

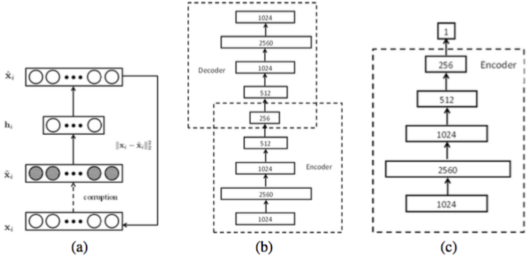

Figure 1: Some key components of the network architecture: (a) denoising autoencoder; (b) stacked denoising autoencoder; (c) network for online tracking.

Offline Training with Auxiliary Data

Loss for a DAE

$$\mathcal{L} = \sum_{i=1}^{k} ||x_i - \hat{x}_i||_2^2 + \lambda(||W||_F^2 + ||W'||_F^2) \qquad (1)$$

Sparsity(use Cross-Entropy)

- ρ_j : the target sparsity level of the j-th unit;
- $\hat{\rho}_j$: average empirical activation of the j-th unit;

$$H(\rho||\hat{\rho}) = -sum_{j=1}^{m}[\rho_j \log(\hat{\rho}_j) + (1-\rho_j)\log(1-\hat{\rho}_j)] \qquad (2)$$

Online Tracking Process

Structure

1. A sigmoid classification layer is then added to the encoder part of the SDAE obtained from offline training.
2. When a new video frame arrives, we first draw particles according to the particle filter approach.
3. The confidence p_i of each particle is then determined by making a simple forward pass through the network.

Online Tracking Process

Threshold(tradeoff)

If the maximum confidence of all particles in a frame is below a predefined threshold τ, it may indicate significant appearance change of the object being tracked. To address this issue, the whole network can be tuned again in case this happens.

DLT Implementation Details

Gradient based method

- momentum : 0.9

SDAE(offline)

- Noise type : Gaussian noise with a variance of 0.0004
- λ : 0.0001, ρ_i : 0.05
- batch size : 100

SDAE(online)

- λ : 0.002(to avoid overfitting)
- batch size :10
- τ : 0.9
- number of particles : 1000