Deep Multiple Instance Learning for Image Classification and Auto-Annotation

`Image annotation/tagging`

Novel ideas

1. **introduce Multiple Instance Learning**
2. **regard the object proposals and text annotations as two instance sets**
3. **DMIL for Image Representations; DNN for Text annotations**
4. **A new Dataset created by hand-crafted for patch-level image annotation**

Model

Full framework for learning regions and keywords

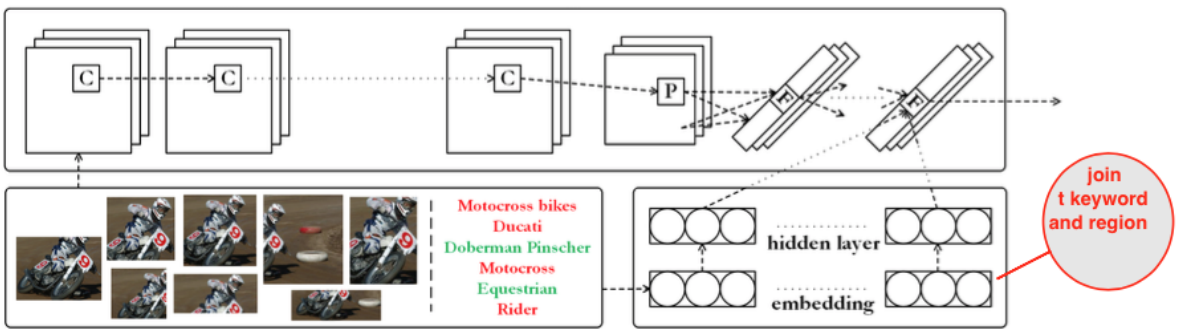
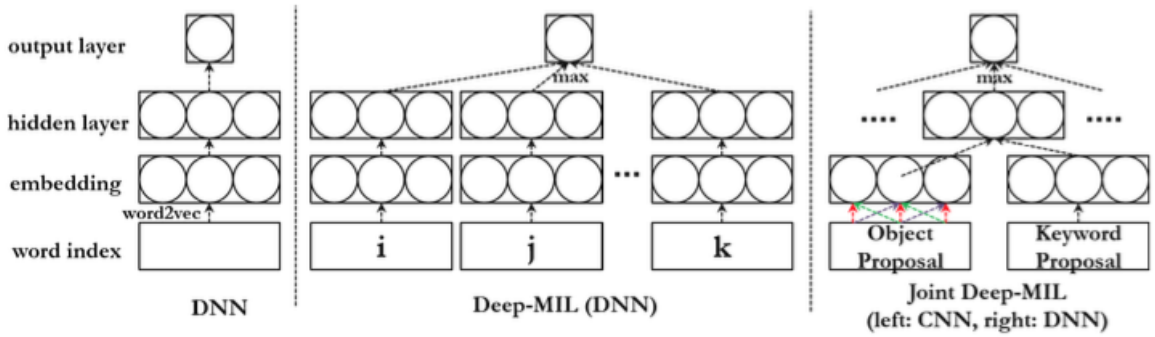
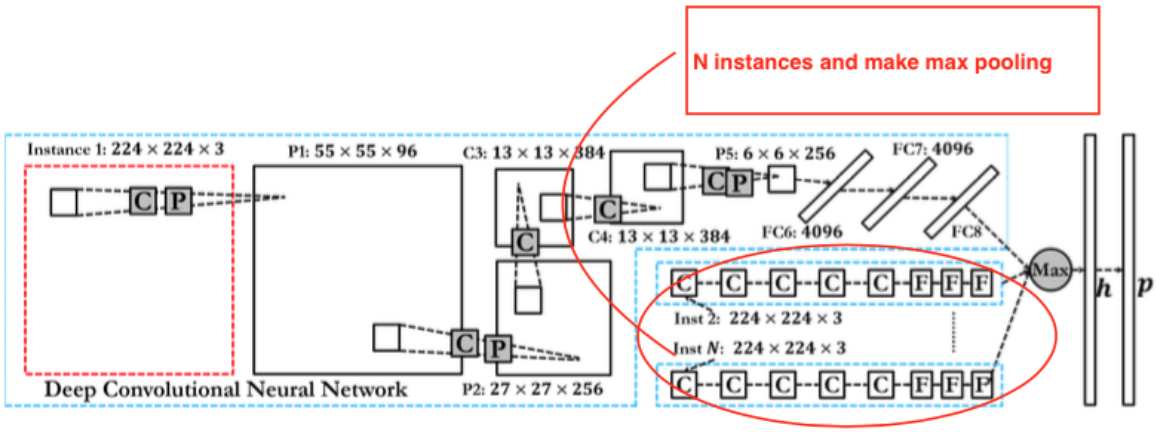


Figure 5. Illustration of our framework for jointly learning image regions and keywords. Here P stands for a pooling layer, C for a convolution layer, and F for a fully connected layer.

Joint Deep-MIL for Image Keywording



Framework for learning deep visual representations with a MIL



Extraction details

1. **Keyword extraction from web data:** search the image from Baidu to find a set of most similar images and surrounding documents.
2. **Word-to-vector feature:** employ a simple DNN contains one input layer, one hidden layer, and one output layer with softmax. Finally, a 128D W2V feature is used to relieve the computational burden.
3. **Combine the outputs of image and text understanding system in the final fully connected layer.**

$$\bar{h}_i = f\begin{pmatrix} h_{i11} & h_{i12} & \dots & h_{i1n} \\ h_{i21} & h_{i22} & \dots & h_{i2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{im1} & h_{im2} & \dots & h_{imn} \end{pmatrix}$$

where m is the number of keywords and n is the number of patches.

Dataset for annotation

- **50 categories, each category have 50 images, manually label bounding boxes**
- **collect keywords while restricting them to be from a dictionary of 981 nouns**

Experiences details

- **Pretrain the model on the ILSVRC dataset, then train framework on the PASCAL 07**
- **Use BING as the proposal generating system, retain the windows with confidence scores larger than 0.97**