Image Tagging

CS771A Course Presentation

Project Supervisor: Professor Piyush Rai

Problem Statement

We worked on the problem of automatic image annotation. Given an image, our task is to place it in a set of categories. The problem can be viewed as a multi-class classification problem, with the tags being the classes.

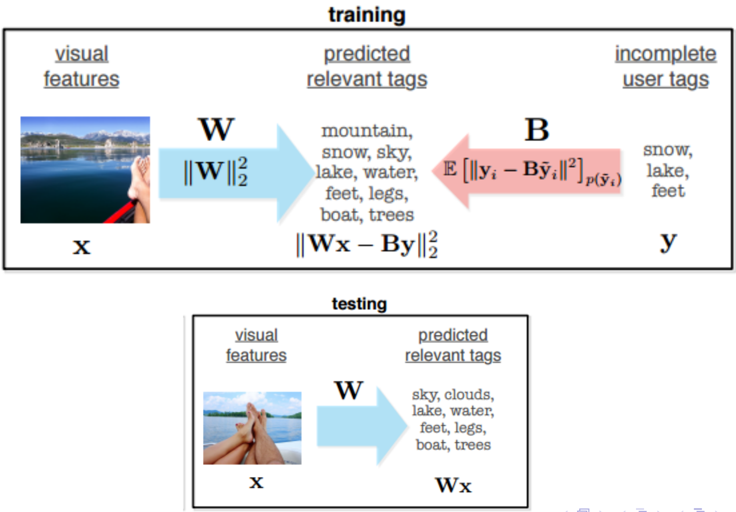
green, wildlife, spring, desert, Nevada

bike, bicycle, glow, sunlight, light, people, summer, august, France, Paris, silhouette, mist, silhouettes, olympus, blue

FASTTAG

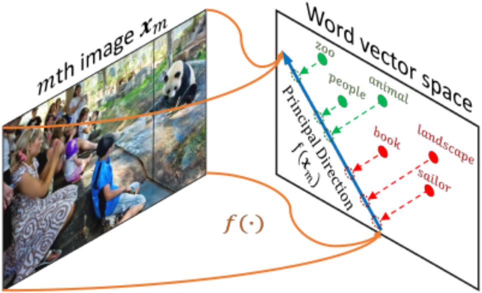
The algorithm interprets training data as images with partial tags and learns two classifiers to predict tag annotations. The paper gives closed form expressions for the calculation of $\boldsymbol{B}$ and $\boldsymbol{W}$.

FASTTAG


training
visual features
predicted relevant tags
incomplete user tags
$\mathbf{W}$
$\|\mathbf{W}\|_2^2$
mountain, snow, sky, lake, water, feet, legs, boat, trees
$\mathbf{B}$
$\mathbb{E}\left[\|\mathbf{y}_i - \mathbf{B}\tilde{\mathbf{y}}_i\|^2\right]_{p(\tilde{\mathbf{y}}_i)}$
snow, lake, feet
$\mathbf{x}$
$\|\mathbf{W}\mathbf{x} - \mathbf{B}\mathbf{y}\|_2^2$
$\mathbf{y}$
testing
visual features
predicted relevant tags
$\mathbf{W}$
sky, clouds, lake, water, feet, legs, boat, trees
$\mathbf{x}$
$\mathbf{W}\mathbf{x}$

FAST0TAG

Learns a principal direction in the word vector space corresponding to each image, such that the positive tags of the image are ranked ahead of all the tags along that direction.



Ranking SVM: Learn the principal direction corresponding to each image using Ranking SVM and then learn a linear mapping between them.
Neural Network Methods: Combine the learning of the principal direction and the mapping into one framework.

Matrix Factorisation

- We try to learn embeddings for each of the tags from the co-occurence matrix.
- Given the co-occurence matrix X we formulate it as a matrix factorisation problem.

$$X = UU^T$$

- We minimize the following loss function.

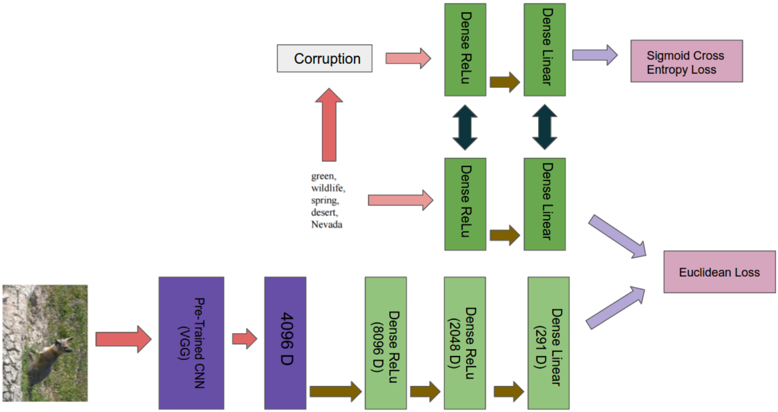
$$L = \sum_{n=1}^{N} \sum_{m=1}^{N} (X_{nm} - u_n^T v_m)^2 + \lambda_u \sum_{n=1}^{N} ||u_n||^2 + \lambda_v \sum_{m=1}^{N} ||v_m||^2 + \lambda_{uv} \sum_{n=1}^{N} ||u_n - v_m||^2$$

- These embeddings are used as a substitute for word vectors.

Kernelized Ridge Regression

- We tried to kernelize the linear mapping from the images to their principal directions.
- We experimented with different kernels and saw significant improvement of results in some cases.

Neural Network Methods: FastTag



Experimental Results

We have used **IAPRTC-12** for all our experiments. The dataset consists of 19,627 images of sports, actions, people, animals, cities, landscapes and many other aspects of contemporary life. Tags are extracted from the free-flowing text captions accompanying each image. Over-all, 291 tags are used.

Algorithm	F1 Score(%)
FastTag	32
Fast0Tag(RankSvm)	30
Fast0Tag(Neural Net)	45
Co-occurrence embeddings concatenated with word vector (Fast0Tag)	42.9
Kernelized Ridge Regression (Fast0Tag)	45

Key Learnings

- We were amazed by the fact that a problem can be solved by completely different and novel approaches.
- Although several standard methods exist for learning a good model, one must not feel restricted to them.
- Application of various concepts taught in the class on real world problems helped us internalize those concepts.

THANK YOU

From **:**
Jayant Agrawal, Rushab Munot, Sagar Chand, Rishabh Goyal and Srishti Jain