# Instagram Hashtag Prediction, With and Without data

**Shreyash Pandey** (shreyash@stanford.edu), **Abhijeet Phatak** (aphatak@stanford.edu)

**Stanford University**

## Objectives

To demonstrate the effectiveness of Zero Shot Learning (ZSL) on an image recognition task by comparing it to the cascade of data collection, data cleaning and fully supervised training.

1. Data collection and cleaning.
2. Supervised learning with less and noisy data.
3. ZSL from well learned visual and textual models.

## Example Output

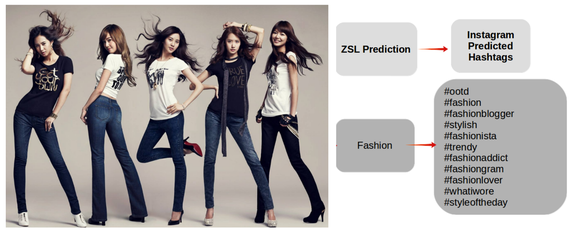



Figure 1: Instagram Hashtag Prediction

## Introduction

1. **Aim** - Instagram hashtag prediction.
2. **Challenge** - No publicly available dataset.
3. **Methods** - 1. Collect data, clean it and train. 2. Try and manage without data (ZSL).
4. **ZSL** is based on knowledge transfer. Uses information about unseen classes from text corpora, mimicking how humans learn.
5. **Word2Vec features** - Rich, informative vectors that capture similarity between English words.
6. For fair **comparison** of the two approaches - keep a mapping of Instagram relevant English words to common hashtags (deterministic mapping)
7. **Outcomes** for fully supervised and ZSL models are expected to be subjectively similar.

## Fully Supervised Prediction

### Data Collection

- Collected 76 common English words (classes) relevant to Instagram.
- Used DuckDuckGo image search to download 200 images per class.

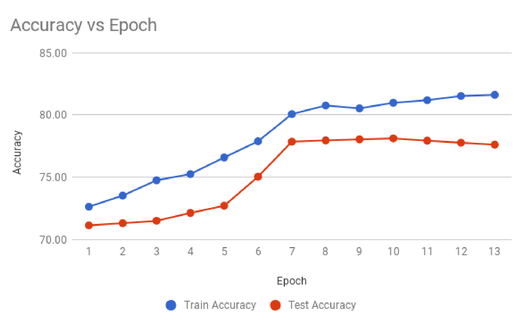
### Data Processing

- Feature extraction using AlexNet FC7 $\rightarrow$ 4096 dimensional feature vectors.
- PCA dimensionality reduction - 4096 $\rightarrow$ 128 dimensions.
- For each class, detect outliers with K-Means on 128D vectors ($2\sigma$).

### Supervised Training

- Split into train and val (80:20)
- Baseline: SVM with bag of words model
- Transfer Learning : Fine-tune a pre-trained ResNet-18 to our dataset (ensures that the network does not overfit and trains quickly).
- 13 epochs of training.

### Learning Curve




## Zero Shot Learning

1. ConSE (Convex Combination of Semantic Embeddings)
   - Pre-trained ImageNet ResNet-18, obtain ImageNet predictions for images of unseen classes
   - Using the probabilities as coefficients, obtain a convex combination of word vectors
   - Predict the unseen class using a nearest neighbor search in the word space.

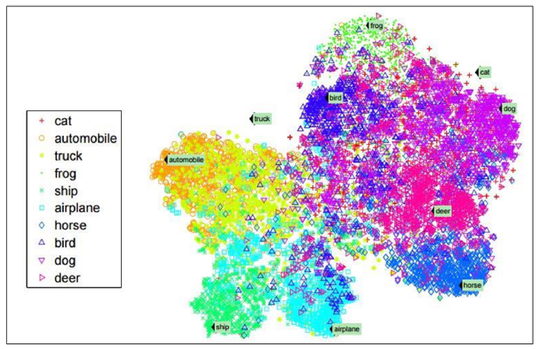



Figure 2: t-SNE visualization of a word space

## Hashtag Prediction

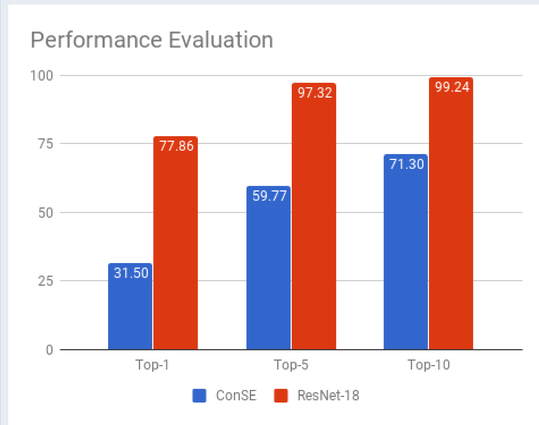
- Map predicted tags to hash-tags.
- For Fully supervised, take top 2 predictions and sample based on probabilities
- For ZSL, take top 5 predictions and sample based on probabilities
- Hashtags generated by ZSL and ResNet-18 are subjectively similar and accurate.

## Implementation Details

- PyTorch and Torchvision Deep Learning Framework for CNNs.
- MATLAB Bag of Visual Words for SVM.
- Gensim for 300 dimensional word vectors trained on Google News with a vocab size of 3M.
- Scrape hash-tags from all-hashtag.com for 76 English words
- Distance metric for KNN $\rightarrow$ *cosine* similarity.

## Analysis

### Comparison of ConSE and ResNet-18




### Results

| Model | Top-1 Accuracy |
|---|---|
| Supervised SVM | 19 |
| Supervised ResNet-18 | 77 |
| ZSL ConSE | 32 |

Table 1: Accuracy Comparison

| Model | Hashtags |
|---|---|
| ResNet-18 | #ootd#fashion#trendy |
| ConSE | #outfit#whatiwore#dress |

Table 2: Hashtags Comparison for Image in Fig. 1

## Future Work

- Compare performance of ResNet-18 with SOTA ZSL classifiers.
- Improve English word to Hashtag mapping.
- Analyze the performance using a confusion matrix to find out which classes are visually and semantically similar.
- Make a website/application that is based on these ideas.