Webly Supervised Learning Meets Zero-shot Learning: A Hybrid Approach for Fine-grained Classification

Li Niu
Fine-grained classification: distinguish the subtle difference between subcategories.
Fine-grained classification: distinguish the subtle difference between subcategories.
Collecting training set for fine-grained classification is difficult due to two reasons.

1. There are a huge number of subcategories within each category.

14,000 bird species in the world

[Welinder et al., 2010]
Collecting training set for fine-grained classification is difficult due to two reasons.

2. Labeling subcategories requires professional knowledge.

<table>
<thead>
<tr>
<th>forehead color</th>
<th>black</th>
</tr>
</thead>
<tbody>
<tr>
<td>breast pattern</td>
<td>solid</td>
</tr>
<tr>
<td>breast color</td>
<td>white</td>
</tr>
<tr>
<td>head pattern</td>
<td>plain</td>
</tr>
<tr>
<td>back color</td>
<td>white</td>
</tr>
<tr>
<td>wing color</td>
<td>grey</td>
</tr>
<tr>
<td>leg color</td>
<td>orange</td>
</tr>
<tr>
<td>size</td>
<td>medium</td>
</tr>
<tr>
<td>bill shape</td>
<td>dagger</td>
</tr>
<tr>
<td>wing shape</td>
<td>pointed</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>primary color</td>
<td>white</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>forehead color</th>
<th>red</th>
</tr>
</thead>
<tbody>
<tr>
<td>breast pattern</td>
<td>solid</td>
</tr>
<tr>
<td>breast color</td>
<td>white</td>
</tr>
<tr>
<td>head pattern</td>
<td>capped</td>
</tr>
<tr>
<td>back color</td>
<td>white/black</td>
</tr>
<tr>
<td>wing color</td>
<td>white/black</td>
</tr>
<tr>
<td>leg color</td>
<td>black</td>
</tr>
<tr>
<td>size</td>
<td>medium</td>
</tr>
<tr>
<td>bill shape</td>
<td>all-purpose</td>
</tr>
<tr>
<td>wing shape</td>
<td>pointed</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>primary color</td>
<td>white/black</td>
</tr>
</tbody>
</table>
Collecting **accurate** training set is difficult, but collecting **inaccurate** training set (e.g., crawl from image website) is much easier.
Collecting accurate training set is difficult, but collecting inaccurate training set (e.g., crawl from image website) is much easier.
There are two major issues when learning from web data.

1. label noise
Learning from Web Data

There are two major issues when learning from web data.

2. domain shift

Domain adaptation [1]: adapt $f^s$ to $f^t$ by addressing the domain shift.

Given an entire set of subcategories (e.g., 14,000 bird species):

1. Collect a labeled subset of subcategories and crawl web images for the other subcategories.  
   labeled subcategories (e.g., 1,000)  web subcategories (e.g., 13,000)

Entire set of subcategories = labeled subcategories + web subcategories
Given an entire set of subcategories (e.g., 14,000 bird species):

1. Collect a labeled subset of subcategories and crawl web images for the other subcategories.
   labeled subcategories (e.g., 1,000)  web subcategories (e.g., 13,000)

   Entire set of subcategories = labeled subcategories + web subcategories

2. Use category-level information to transfer knowledge from labeled subcategories to web subcategories, to handle the label noise and domain shift of web data.

   Category-level information can be obtained based on free web data.
Category-level Information

Textual feature of wiki page

Word vector of category name [2]

Linguistic model

vector(king) - vector(man) + vector(woman) ≈ vector(queen)

Our Flowchart

labeled subcategories

well-labeled images

Chihuahua
Shih-Tzu

category: dog

visual features

Our learning model

web subcategories

crawl web images

Briard
Cardigan

visual features
Our Flowchart

labeled subcategories

well-labeled images

subcategory names

linguistic model

visual features

word vector

visual features

Our learning model

category: dog

Chihuahua

Shih-Tzu

…

Briard

Cardigan

…

crawl web images

labeled subcategories

web subcategories

subcategory names

linguistic model

visual features

word vector

visual features

Our learning model
Our Flowchart

- **labeled subcategories**
  - well-labeled images
  - Chihuahua
  - Shih-Tzu
  - ... subcategory names
  - linguistic model
  - visual features

- **web subcategories**
  - crawl web images
  - Briard
  - Cardigan
  - ... subcategory names
  - word vector
  - visual features

- **unlabeled test images**

**category:** dog

**Our learning model**
Our Flowchart

Our learning model
Our Learning Model

Dictionary learning based method

**Step1:** Learn knowledge (dictionary $D^l$) on labeled subcategories.

$$\min_{D^l} \|X^l - D^l A^l\|_F^2 + \|D^l\|_F^2$$

- $X$: visual features
- $D$: dictionary
- $A$: word vectors
Dictionary learning based method

**Step2:**
- **Transfer knowledge** from labeled subcategories to web subcategories.
- **Handle the label noise** and **domain shift** for web data.

Our Learning Model

\[
\min_{D^w, A^t, \theta} \left\| X^t - D^w A^t \right\|_F^2 + \left\| D^w - D^t \right\|_F^2 + \left\| A^t \right\|_* + \frac{1}{n^w} \left\| X^w \theta - \frac{1}{n^t} X^t 1 \right\|_2^2 + \left\| (X^w - D^w A^w) \Theta \right\|_{2,1}^1
\]

**knowledge transfer**

**domain shift**

s.t. \(1'\theta = n^w, \ 0 \leq \theta \leq b1\)

θ : assign different weights on different web training images, identify the non-outliers with the closest data distribution to test data.
Our Learning Model

Dictionary learning based method

Step 2: ■ Transfer knowledge from labeled subcategories to web subcategories.
■ Handle the label noise and domain shift for web data.

$$\min_{D^w, A^t, \theta} \|X^t - D^w A^t\|_F^2 + \|D^w - D^l\|_F^2 + \|A^t\|_*$$

\[+ \|\frac{1}{n_w} X^w \theta - \frac{1}{n_t} X^t 1\|_2^2 + \|(X^w - D^w A^w) \Theta\|_{2,1}\]

s.t. \[1' \theta = n_w, \quad 0 \leq \theta \leq b1\]

\[\theta:\] assign different weights on different web training images,
identify the non-outliers with the closest data distribution to test data.
### Fine-grained classification

CUB, SUN, Dogs + Flickr images (LS)

CUB, SUN, Dogs (WS)

<table>
<thead>
<tr>
<th></th>
<th>CUB</th>
<th>SUN</th>
<th>Dogs</th>
</tr>
</thead>
<tbody>
<tr>
<td># Labeled Subcategories (LS)</td>
<td>150</td>
<td>707</td>
<td>85</td>
</tr>
<tr>
<td># Web Subcategories (WS)</td>
<td>50</td>
<td>10</td>
<td>28</td>
</tr>
</tbody>
</table>

#### Experiments

- **Only use WS**
- **Only use LS**
- **Naive combination of WS and LS**
- **Our method**
Identified outliers and non-outliers based on our learnt weights $\theta$

<table>
<thead>
<tr>
<th>Identified outliers</th>
<th>Non-outliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75</td>
<td>1.46</td>
</tr>
<tr>
<td>0.75</td>
<td>1.46</td>
</tr>
</tbody>
</table>
Thank you!