Deep Self-Learning From Noisy Labels

Presented by: Ashkan Farhangi

Jiangfan Han
Ping Luo
Xiaogang Wang
Outline

• Introduction
• Related works
• Proposed method
• Experiments & results
History: ImageNet

- Large-scale well-labeled training data.
- 14 million images hand-annotated by the project.
- Neural Networks achieve better performance than humans.
- Quite expensive data labeling process.

http://www.image-net.org/
Solution

• Supervised and unsupervised deep learning algorithms
Solution

• Automatically collect noisy label data from Internet.
Noisy label

Clean cat

Noisy cat

Clean jasmine

Noisy jasmine
Noisy label challenges

• Researchers need to develop algorithms for learning in presence of label noise.
• Human supervision for label correction is costly but effective.
• Approaches not relying on human supervision are scalable but less effective.
• There is a need to automate the label correction process.
Distribution of features

Existing solutions

• **Supervised**: Supervised methods learn a binary classification from verification labels for each class.
  • CleanNet, 2-layer MLP, kNN, SVM

• **Unsupervised**: Unsupervised methods rely on the outlier removal.
  • DRAE

• **Weakly supervised**: verification labels are not required.
  • Classification filtering, Loss Factorization
Related Work – CleanNet

• Utilizes two networks in one framework for end-to-end learning of image classifiers with label noise
• Uses CNN- image classifier to provides more discriminative features.
• It also uses attention mechanism to assign weights to data samples
Related Work – CleanNet

• CleanNet uses the concept of class prototype to represent each class category
Related Work – CleanNet

• Class prototypes can effectively represent classes.
• Compare the sample image with the prototype will determine if the label is correct.
Related Work – CleanNet

Correct label

Noisy label

Related work

• Majority of these approaches require extra clean samples.
• Using a complicated training procedure.
• Relying on the noisy dataset that is created intentionally by human with label flipping.
  • CIFAR10
Deep Self-Learning for noisy labels

• Select images as prototypes by their density and similarity according to the data distribution.
• Not rely on clean supervision images.
• Optimal number of prototypes for each class.
Deep Self-Learning for noisy labels
Proposed network
Training Phase

\[ \mathcal{L}(\mathcal{F}(\theta, x), y) = -\frac{1}{n} \sum_{i=1}^{n} \log \left( \mathcal{F}(\theta, x_i)y_i \right) \]
Training Phase Losses

\[ \mathcal{L}(\mathcal{F}(\theta, x), y) = -\frac{1}{n} \sum_{i=1}^{n} \log(\mathcal{F}(\theta, x_i)_{y_i}) \]

\[ \mathcal{L}_{\text{total}} = (1 - \alpha)\mathcal{L}(\mathcal{F}(\theta, x), y) + \alpha\mathcal{L}(\mathcal{F}(\theta, x), \hat{y}) \]
Label Correction Phase

$$\sigma_c = \frac{1}{p} \sum_{l=1}^{p} \cos(\mathcal{G}(x), \mathcal{G}(x_{cl})), c = 1...K$$

$$\hat{y} = \arg\max_c \sigma_c, \ c = 1...K$$
Proposed network

\[ \mathcal{L}_{\text{total}} = (1 - \alpha)\mathcal{L}(\mathcal{F}(\theta, x), y) + \alpha\mathcal{L}(\mathcal{F}(\theta, x), \hat{y}) \]
Distribution

- Over 80% of the samples have $\eta > 0.9$
- Half of the samples have $\eta > 0.95$.
- high-density value $\rho$ and low similarity value $\eta$ can be chosen as a prototype.
Food101-N Dataset

• The Food-101N dataset was introduced in the CleanNet paper.
• 310,009 images of food recipes classified in 101 classes (categories).

Source: https://kuanghuei.github.io/Food-101N/
Clothing1M Dataset

- It consists of 1M images with noisy class labels from 14 fashion classes
# Experiment results

<table>
<thead>
<tr>
<th>#</th>
<th>Method</th>
<th>Data</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cross Entropy</td>
<td>1M noisy</td>
<td>69.54</td>
</tr>
<tr>
<td>2</td>
<td>Forward [25]</td>
<td>1M noisy</td>
<td>69.84</td>
</tr>
<tr>
<td>3</td>
<td>Joint Optim. [35]</td>
<td>1M noisy</td>
<td>72.23</td>
</tr>
<tr>
<td>4</td>
<td>MLNT-Teacher [16]</td>
<td>1M noisy</td>
<td>73.47</td>
</tr>
<tr>
<td>5</td>
<td>Ours</td>
<td>1M noisy</td>
<td><strong>74.45</strong></td>
</tr>
<tr>
<td>6</td>
<td>Forward [25]</td>
<td>1M noisy + 25k verify</td>
<td>73.11</td>
</tr>
<tr>
<td>7</td>
<td>CleanNet $w_{hard}$ [15]</td>
<td>1M noisy + 25k verify</td>
<td>74.15</td>
</tr>
<tr>
<td>8</td>
<td>CleanNet $w_{soft}$ [15]</td>
<td>1M noisy + 25k verify</td>
<td>74.69</td>
</tr>
<tr>
<td>9</td>
<td>Ours</td>
<td>1M noisy + 25k verify</td>
<td><strong>76.44</strong></td>
</tr>
<tr>
<td>10</td>
<td>Cross Entropy</td>
<td>1M noisy + 50k clean</td>
<td>80.27</td>
</tr>
<tr>
<td>11</td>
<td>Forward [25]</td>
<td>1M noisy + 50k clean</td>
<td>80.38</td>
</tr>
<tr>
<td>12</td>
<td>CleanNet $w_{soft}$ [15]</td>
<td>1M noisy + 50k clean</td>
<td>79.90</td>
</tr>
<tr>
<td>13</td>
<td>Ours</td>
<td>1M noisy + 50k clean</td>
<td><strong>81.16</strong></td>
</tr>
</tbody>
</table>

Clothing1M

<table>
<thead>
<tr>
<th>#</th>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cross Entropy</td>
<td>84.51</td>
</tr>
<tr>
<td>2</td>
<td>CleanNet $w_{hard}$ [15]</td>
<td>83.47</td>
</tr>
<tr>
<td>3</td>
<td>CleanNet $w_{soft}$ [15]</td>
<td>83.95</td>
</tr>
<tr>
<td>4</td>
<td>Ours</td>
<td><strong>85.11</strong></td>
</tr>
</tbody>
</table>
Prototypes per class

• The number of class prototypes is the key to the representation ability to a class

\[ L_{\text{total}} = (1 - \alpha)L(\mathcal{F}(\theta, x), y) + \alpha L(\mathcal{F}(\theta, x), \hat{y}) \]
Weight factor

• Decides the network will concentration on the noisy labels and corrected labels.
Deep self-learning method – Food101-N

Edamame Dumpling Apple pie Cup cake Churros Ice cream Sashimi Sushi
Deep self-learning method – Clothing1M
Insights from hard classes
Cluster methods

• The classification accuracy with various clustering methods
• Further research can be done in this area to find prototypes.

<table>
<thead>
<tr>
<th>Method</th>
<th>Data</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-means++ [1]</td>
<td>1M noisy</td>
<td>74.08</td>
</tr>
<tr>
<td>Density peak Euc. [31]</td>
<td>1M noisy</td>
<td>74.11</td>
</tr>
<tr>
<td>Ours</td>
<td>1M noisy</td>
<td><strong>74.45</strong></td>
</tr>
<tr>
<td>K-means++ [1]</td>
<td>1M noisy + 25k verify</td>
<td>76.22</td>
</tr>
<tr>
<td>Density peak Euc. [31]</td>
<td>1M noisy + 25k verify</td>
<td>76.05</td>
</tr>
<tr>
<td>Ours</td>
<td>1M noisy + 25k verify</td>
<td><strong>76.44</strong></td>
</tr>
</tbody>
</table>
Sources:


Thank you.

Questions?