DADA: Deep Adversarial Data Augmentation for Extremely Low Data Regime Classification

Xiaofeng Zhang\textsuperscript{1}, Zhangyang Wang\textsuperscript{2}, Dong Liu\textsuperscript{1}, Qing Ling\textsuperscript{3}

Presenter: Weiyu Li\textsuperscript{1}

\textsuperscript{1}University of Science and Technology of China
\textsuperscript{2}Texas A&M University
\textsuperscript{3}Sun Yat-Sen University

May 14, 2019
1. Introduction

2. Related Work

3. Technical Approach

4. Simulations

5. Experiments with Real-world Small Data

6. Conclusion
Deep learning has revolutionary performance given sufficient labeled data, but overfits easily and generalizes poorly when data is insufficient.

Focus: training a deep neural network with extremely low data regime.

In reality, either massively annotating labels is a labor-intensive task, or only limited datasets are available.

We focus on this area.
Related Work

Deep learning on Small Samples

- **Data Augmentation**
  - Artificially synthesizing new labeled samples from existing ones
  - Image rotation, lighting/color tone modification, rescaling and so on

- **Pre-training and Semi-Supervised Learning**
  - Utilizing extra unlabeled data from the same training distribution
  - Heavily relying on the abundance of unlabeled data

- **Transfer Learning**
  - Using unlabeled data from source domain with similar distribution
  - Limited when source and target domains possess notable discrepancy
Related Work
Deep learning on Small Samples

- We place ourself in front of an even more ill-posed extremely low data regimes: only a small set of labeled data are available, and nothing - including unlabeled data - else
- Generative adversarial network has gathered a significant amount of attention due to its ability to learn generative models of multiple natural image datasets
Related Work

Generative Adversarial Networks

Figure: The architectures of Vanilla GAN (left), Conditional GAN (mid) and Improved GAN (right)
3 Technical Approach
Technical Approach
Problem Formulation and Solution Overview

Figure: An illustration of DADA architecture
Technical Approach

Going More Discriminative: From \( k + 1 \) Loss to \( 2k \) Loss

DADA: a more “economical” method

- The labeled data should benefit classifier learning more than unlabeled ones
- The generated labeled samples should join force with the real labeled samples, and their decision boundary should be well aligned

We design our DADA architecture and build \( 2k \) loss function
**Technical Approach**

Going More Discriminative: From $k + 1$ Loss to $2k$ Loss

Table: The Comparison of Loss Functions among GAN Discriminators

<table>
<thead>
<tr>
<th>Model</th>
<th>Class Number</th>
<th>Classes</th>
<th>Training Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vanilla GAN</td>
<td>2</td>
<td>real, fake</td>
<td>unlabeled only</td>
</tr>
<tr>
<td>Conditional GAN</td>
<td>2</td>
<td>real, fake</td>
<td>labeled only</td>
</tr>
<tr>
<td>Improved GAN</td>
<td>$k + 1$</td>
<td>$real_1, \ldots, real_k; fake$</td>
<td>labeled + unlabeled</td>
</tr>
<tr>
<td>DADA</td>
<td>$2k$</td>
<td>$real_1, \ldots, real_k; fake_1, \ldots, fake_k$</td>
<td>labeled only</td>
</tr>
</tbody>
</table>
The training procedure of DADA is divided into two different phases.

**Generation Training**

The classifier and the augmenter compete with each other within a specific class. The game between the two players will have its optimum only if \( p_{data}(x|y) = p_g(x|y) \).

\[
L_C^I = - \mathbb{E}_{x,y \sim p_{data}(x,y)} \left[ \log p(y|x, y < k + 1) \right] \\
- \mathbb{E}_{x,y \sim p_g(x,y)} \left[ \log p(y|x, k < y < 2k + 1) \right]
\]

\[
L_A^I = - \mathbb{E}_{x,y \sim p_g(x,y)} \left[ \log p(y - k|x, k < y < 2k + 1) \right] \\
+ \lambda \| \mathbb{E}_{x,y \sim p_{data}(x,y)} f(x|y) - \mathbb{E}_{z \sim p_z(z), y \sim p_c} f(G(z, y)|y) \|
\]
The training procedure of DADA is divided into two different phases.

**Classification Training**

The augmentor is fixed just as a data provider. We only train the classifier on the generated data and real data.

\[
\mathcal{L}_{C}^{II} = \mathcal{L}_{data} + \mathcal{L}_{gen}
\]

\[
\mathcal{L}_{data} = - \mathbb{E}_{x,y \sim p_{data}(x,y)} [\log p(y|x, y < k + 1) \\
+ p(y + k|x, y < k + 1)]
\]

\[
\mathcal{L}_{gen} = - \mathbb{E}_{x,y \sim p_{g}(x,y)} [\log p(y|x, k < y < 2k + 1) \\
+ p(y - k|x, k < y < 2k + 1)]
\]
1 Introduction

2 Related Work

3 Technical Approach

4 Simulations

5 Experiments with Real-world Small Data

6 Conclusion
We intentionally sample the given training data to simulate the extremely low data regimes, and compare the following training options:

- **C**: directly train a classifier using the limited training data
- **C\_augmented**: perform traditional data augmentation
- **DADA**: the proposed data augmentation
- **DADA\_augmented**: first apply traditional augmentation, then perform DADA
- **Vanilla GAN** (which adopts 2-class loss)
- **Improved-GAN** (which adopts \((k + 1)\)-class loss)
Simulation
CIFAR-10 Dataset

Figure: Results on CIFAR-10, the test accuracy in different training settings with respect to the number of training images per class.
1. Introduction

2. Related Work

3. Technical Approach

4. Simulations

5. Experiments with Real-world Small Data

6. Conclusion
CIBS-DDSM: A medical labeled dataset for tumor classification (benign or malignant)

We compare DADA with existing learning-based data augmentation work Tanda (Ratner et al. 2017) on CIBS-DDSM dataset in which the effects of transfer learning are limited.

Table: Comparison between DADA and Tanda.

<table>
<thead>
<tr>
<th>Models</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanda (MF)</td>
<td>0.5990</td>
</tr>
<tr>
<td>Tanda (LSTM)</td>
<td>0.6270</td>
</tr>
<tr>
<td>DADA</td>
<td>0.6196</td>
</tr>
<tr>
<td>DADA_augmented</td>
<td>0.6549</td>
</tr>
</tbody>
</table>
Content

1 Introduction

2 Related Work

3 Technical Approach

4 Simulations

5 Experiments with Real-world Small Data

6 Conclusion
Conclusions

- DADA: a learning-based data augmentation solution for training deep classifiers in extremely low data regimes
- GAN with a new $2k$ loss: enhancement of diversity
- Simulations and real-data experiments: the practical advantage of DADA
- Future work: object detection
Thank you!