DeepSMOTE: Deep Learning For Imbalanced Data

UNIVERSITY OF NOTRE DAME | RESEARCH

VCU

Damien Dablain[1], Bartosz Krawczyk[2], Nitesh Chawla[1]

[1]Department of Computer Science and Engineering, University of Notre Dame, IN, USA
[2]Department of Computer Science, Virginia Commonwealth University, Richmond, VA USA

Abstract

Imbalanced data is a significant challenge for modern deep learning systems.

The two main approaches to imbalanced data are: **modifying loss functions** and **resampling,** which suffer from limitations.

Therefore, there is a need for a novel oversampling method that is **specifically tailored to deep learning**, can work on **raw images** and is **capable of generating high-quality output**.

We present ***DeepSMOTE***, which generates information-rich images without the need for a discriminator.

DeepSMOTE Algorithm

Algorithm 1: DEEPSMOTE

Data: B: batches of imbalanced training data $B = \{b_1, b_2, \ldots, b_n\}$

Input: Model parameters: $\Theta = \{\Theta_0, \Theta_1, \ldots, \Theta_j\}$; Learning Rate: α

Output: Balanced training set.

Train the Encoder / Decoder:

for $e \leftarrow epochs$ **do**

  for $m \leftarrow B$ **do**

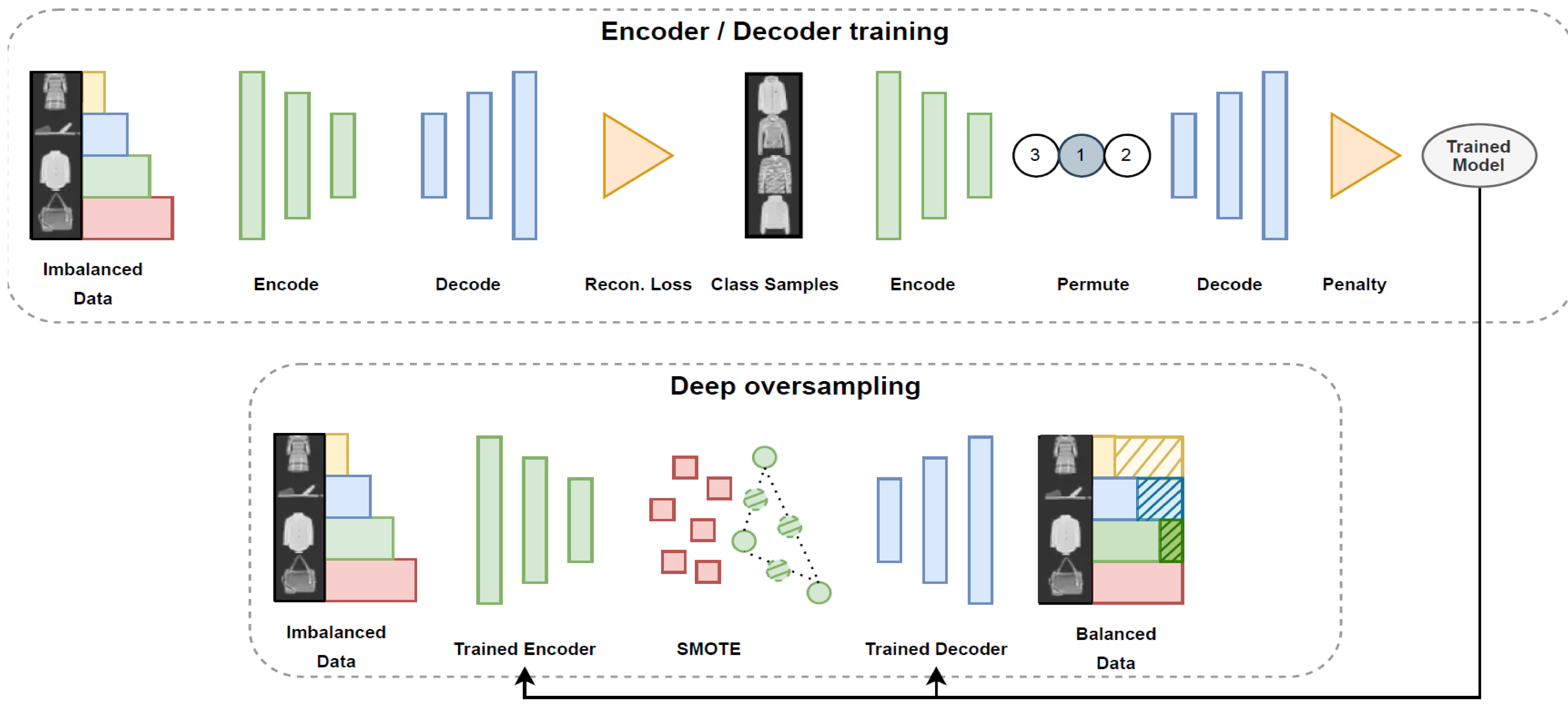
$E_B \leftarrow encode(B)$
$D_B \leftarrow decode(E_B)$
$R_L = \frac{1}{n}\sum_{i=1}^{n}(D_Bi - B_i)^2$
$C_D \leftarrow sample(class\ data)$
$E_S \leftarrow encode(C_D)$
$P_E \leftarrow permute - order - of(E_S)$
$D_P \leftarrow decode(P_E)$
$P_L = \frac{1}{n}\sum_{i=1}^{n}(D_Pi - C_{Di})^2$
$T_L = R_L + P_L$
$\Theta := \Theta - \alpha\frac{\partial T_L}{\partial \Theta}$

Generate Samples:

for $i \leftarrow no.\ of\ minority\ classes$ **do**

$C \leftarrow select(class\ data)$
$E \leftarrow encode(C)$
$G \leftarrow SMOTE(E)$
$S \leftarrow decode(G)$

DeepSMOTE Design



DeepSMOTE has a simple, yet effective design:

- An encoder / decoder framework,
- SMOTE-based oversampling, and
- A dedicated loss function enhanced by a penalty term.

Training

DeepSMOTE is trained in an end-to-end fashion, **without a discriminator**.

During training, an imbalanced dataset is input to the encoder / decoder.

A reconstruction loss is computed based on batched data (which includes both majority and minority class examples).

Next, samples are drawn from the same class. A penalty loss is computed based on a permutation of the order of the samples (e.g., encode D_0, D_1, D_2…and decode D_1, D_2, D_0).

The penalty loss is based on the MSE difference between D_0 and D_1, D_1 and D_2, etc., as if an image (I_0) was oversampled by SMOTE (I_1; i.e., a difference were calculated from the image's nearest neighbor). This step is designed to **insert variance into the encoding / decoding process**.

Experimental Results

	MNIST			*FMNIST*			*CIFAR*			*SVHN*			*CELEBA*		
	ACSA	GM	F1	ACSA	GM	F1	ACSA	GM	F1	ACSA	GM	F1	ACSA	GM	F1
SMOTE	81.48	83.99	82.44	67.94	74.84	67.12	28.02	50.08	29.58	70.18	76.33	71.80	60.29	70.48	60.03
AMDO	84.29	88.73	84.88	74.90	80.89	75.39	31.19	53.99	32.44	71.94	78.52	73.06	63.54	72.86	62.94
MC-CCR	86.19	92.04	86.46	78.58	86.17	79.03	32.83	56.68	33.91	72.01	80.94	74.26	65.23	77.14	64.88
MC-RBO	87.25	94.46	88.69	80.06	88.02	80.14	33.01	59.15	35.83	74.20	82.97	74.91	67.11	80.52	65.37
BAGAN	92.56	96.11	93.85	82.50	90.51	82.96	42.41	64.12	43.01	75.81	86.44	77.02	68.62	80.84	**68.33**
GAMO	95.45	97.61	95.11	83.05	90.76	83.00	44.72	65.72	**45.93**	75.07	86.00	76.68	66.06	79.11	64.85
DeepSMOTE	**96.16**	**98.11**	**96.44**	**84.88**	**91.63**	**83.79**	**45.26**	**66.13**	44.86	**79.59**	**88.67**	**80.71**	**72.40**	**82.91**	66.99

- 5 popular datasets, using 5-fold cross validation, against 4 pixel-based oversampling and two GAN-based methods.
- **Pixel-based methods performed much worse than deep learning approaches**. Only the MC-RBO method was able to deliver results not far from GANs.
- Although GAN-based methods performed better, **DeepSMOTE performed strongly against all 6 methods**.
- We can also see that DeepSMOTE generated **high-quality images.**

Summary

We propose **DeepSMOTE**, which marries the **simplicity** of **shallow learning** (by incorporating SMOTE and metric learning) **with deep architectures** that work on complex data.

DeepSMOTE works in an end-to-end fashion, and generates high quality images that can be used for data augmentation and overcoming class imbalance.

Furthermore, it generates images without the need for a complex discriminator network, which is commonly used by GAN-based oversampling methods.