# SPL: EXPLOITING UNLABELED DATA FOR MULTI-LABEL IMAGE CLASSIFICATION

*Weibo Zhang*[1,2], *Fuqing Zhu*[*,1], *Jiao Dai*[1], *Songlin Hu*[1], *Jizhong Han*[1], *Tao Guo*[1]

1. Institute of Information Engineering, Chinese Academy of Sciences, Beijing, 100093, China
2. School of Cyber Security, University of Chinese Academy of Sciences, Beijing, 100049, China

## ABSTRACT

The utilization of the unlabeled data provides a beneficial attempt for improving the generalization ability of the convolutional neural network (CNN) model, just as what is applied in person re-identification task. Different from that, multi-label image classification aims to predict multiple labels for each given image. The unlabeled data should be properly assigned multiple labels for regularizing the training process of CNN model. To make full use of the unlabeled data, this paper proposes a soft pseudo labeling (SPL) method for multi-label image classification. Specifically, the unlabeled samples are first generated by DCGAN and WGAN-GP. Then, the virtual multiple labels of the generated unlabeled samples are assigned based on an initial confidence value by SoftMax function. Finally, both the generated samples and original training samples are fed into the network as input, in order to learn a CNN model with stronger generalization ability. On three public multi-label image classification datasets (*i.e.*, WIDER-Attribute, NUS-WIDE and MS-COCO), SPL provides a stable improvement over the baseline and produces a competitive performance compared with some existing multi-label image classification methods.

***Index Terms***— Multi-label image classification, unlabeled data, soft pseudo labeling, convolutional neural network

## 1. INTRODUCTION

Multi-label image classification [1, 2, 3] is a visual understanding task for discovering each concept in images. Given a typical image, it aims to assign multiple labels from pre-defined classes. The rich semantic information of an image could be represented as various categories of objects/scenes/attributes by learning a discriminative classification model. Multi-label image classification receives increasing attention and has been widely applied in areas of scene classification [4], video annotation [5] or disease diagnosis [6, 7]. In the future, it will provide technical support of product classification for unmanned convenience stores.

In recent years, the CNN-based [8, 9] deep learning models have prevailed over traditional methods since Krizhevsky *et al.* proposed AlexNet for single-label image classification [8]. Then, this mainstream framework is extended to the community of multi-label image classification [10, 11] for higher accuracy. In this paper, we follow this unified end-to-end framework for multi-label image classification.

A common influencing factor on learning-based task is sufficient data, *i.e.*, the generalization ability of the learned model is usually proportional to the number of properly annotated samples. Data augmentation is an intuitive and significant way of producing adaptive advantages during training process in diverse environment, which has been implemented in other computer vision communities, *e.g.*, person re-identification [12, 13]. Zheng *et al.* [12] introduce the unlabeled data generated by DCGAN [14] as outliers to regularize the original supervised pipeline for improving the discriminative ability of the learned CNN model. Zhong *et al.* [13] employ CycleGAN [15] to settle the camera style adaption problem. The camera style-transferred images with explicit labels are generated to extend the original training set. This paper focuses on multi-label image classification, a task in which the learned CNN model is utilized to predict multi-label results directly instead of single-label pedestrian prediction in person re-identification task. Our work provides possible insights on how to improve the generalization ability of the existing CNN model for multi-label image classification.

One of the candidate solutions of generating the additional unlabeled data is employing generative adversarial network (GAN) [16], which has achieved rapid development in recent years and been applied in many fields, *e.g.*, image generation [17], image super-resolution [18], and pose morphing [19]. Due to the strong data fitting ability of GAN, the generated unlabeled samples are suitable for data augmentation and play a complementary role with the original training samples effectively. We focus on a stable way to generate unlabeled samples rather than exquisite samples construction. Moreover, this paper designs a label assignment to produce virtual multiple labels for the unlabeled data.

Given the aforementioned considerations, this paper proposes a soft pseudo labeling (SPL) method which assigns cor-

[*] *Corresponding author: Fuqing Zhu. E-mail: zhufuqing@iie.ac.cn*

This research is supported in part by the National Key Research and Development Program of China under Grant 2017YFB1010000, the Beijing Municipal Science and Technology Project under Grant Z181100002718004, and the National Natural Science Foundation of China under Grant 61702500 and 61702502.

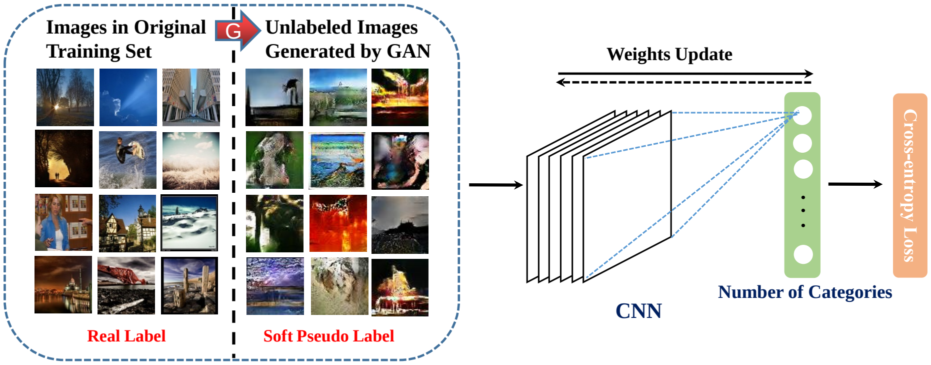

**Fig. 1**. The pipeline of the proposed method for multi-label image classification.

responding virtual multiple labels to the generated unlabeled data, following a semi-supervised learning mechanism for multi-label image classification. Specifically, DCGAN [14] and WGAN-GP [17] are utilized to generate unlabeled samples due to the convenience and stability, respectively. Then, a label assignment method is designed based on an initial confidence value with SoftMax function. While several ways of unlabeled data labeling methods are also discussed, *e.g.*, random labeling, pseudo labeling and smooth labeling. The proposed method has two significant characteristics: *i)* makes full use of the unlabeled data generated by GAN. *ii)* is a scalable pattern which can be integrated into other CNN-based frameworks.

The main contributions of this paper are summarized as follows.

- We make an attempt to introduce GAN as the unlabeled data generator to multi-label image classification task.
- We propose a SPL label assignment method for the generated unlabeled data, combining with the original training samples during the training process of the CNN model. The generalization ability of the learned CNN model is significantly improved.
- On three public widely-used datasets (*i.e.*, WIDER-Attribute [20], NUS-WIDE [21] and MS-COCO [22]), SPL provides a stable improvement over the baseline and produces a competitive performance compared with some existing multi-label image classification methods.

## 2. METHODOLOGY

This paper proposes a SPL label assignment method which assigns virtual labels to the generated unlabeled images, following a semi-supervised learning mechanism for multi-label image classification. Fig. 1 illustrates the pipeline of SPL. In the following subsections, we first explain the problem definition. Then, we briefly describe the unlabeled data generation based on the original training data, which is implemented by DCGAN and WGAN-GP, respectively. Next, the label assignment (*Note:* including some other existing label assignment methods) will be discussed. Finally, we will show the training and testing processes of SPL.

### 2.1. Problem Definition

Let $\mathbf{x}_i$ denote an input image with ground-truth label $\mathbf{y}_i = [y_i^1, y_i^2, \cdots, y_i^c, \cdots, y_i^C]^T$, where $y_i^c$ is a binary indicator. $y_i^c = 1$ if image $\mathbf{x}_i$ is tagged with the $c$-th label, and $y_i^c = 0$ otherwise. $C$ is the total number of all possible labels in the dataset. During training process, the training data is given as $\mathbf{D} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$, where the number of images is $N$. Formally, multi-label image classification problem can be formulated as learning an optimal CNN model $\mathbb{M}$ that maps the input image $\mathbf{x}_i$ to its ground-truth label $\mathbf{y}_i$. During testing process, the label of any testing image $\mathbf{x}_j$ could be predicted through the trained CNN model $\mathbb{M}$, obtaining the final predicted result $\hat{\mathbf{y}}_j$ after an activation function and a threshold.

### 2.2. Unlabeled Data Generation

In this paper, unlabeled data is generated by GAN. GAN is composed of generator and discriminator. Generator generates brand new images and intends to capture the distribution of real data, while discriminator tries to discriminate the generated samples from real data. The images in ordinary training set are fed into the network as input, where DCGAN [14] and WGAN-GP [17] are utilized. Specifically, for DCGAN, the generator is composed of one fully-connected layer followed by four deconvolutional layers as up-sampling function with batch normalization and Tanh functions. The discriminator which consists of several convolutional layers to down-sample the images takes samples from both real data and generator to identify their sources. For WGAN-GP, the generator consists of several deconvolutional layers combined with batch normalization layers and Tanh function as similar as DCGAN. But the discriminator consists of several down-sampling layers combined with layer normalization.

### 2.3. Label Assignment for the Unlabeled Data

In this subsection, we will discuss the label assignment for the generated unlabeled data. Besides the proposed SPL, we first show some hypothetical label assignment proposals, *i.e.*, random labeling, pseudo labeling and smooth labeling.

**Random labeling.** The appearance differences between the original images and generated images by GAN are obvious. Besides, there is almost no specific object in the generated images. In this situation, the most direct assignment way is assigning the random labels for the generated unlabeled data. On multi-label image datasets, there is about 3-class labels per image on average. So we assign a multi-label with random 3-class for each generated image. For a generated image $\mathbf{g}_m$, the label is denoted as $\mathbf{y}_m = [y_m^1, y_m^2, \cdots, y_m^c, \cdots, y_m^C]^T$, where $C$ is the dimension of multi-label, $c = 1, 2, \cdots, C$, and $y_m^c \in \{0, 1\}$. The random multi-label $\mathbf{y}_m$ of the generated image should satisfy $\sum\limits_{c=1}^{C} y_m^c = 3$.

**Pseudo labeling.** Lee *et al.* [23] propose a semi-supervised learning method which automatically assigns a pseudo label for each unlabeled data. First, a basic model is learned by the original labeled training data. Second, unlabeled data is passed through the learned model. The confidence value $\mathbf{v}_m = [v_m^1, v_m^2, \cdots, v_m^c, \cdots, v_m^C]^T$ of unlabeled data for each category is obtained, where $C$ is the dimension of multi-label, $c = 1, 2, \cdots, C$. Third, the top-$K$ highest confidence values are assigned as valid multiple labels for unlabeled data. The pseudo multi-label $\mathbf{y}_m = [y_m^1, y_m^2, \cdots, y_m^c, \cdots, y_m^C]^T$ of the generated image $\mathbf{g}_m$ should satisfy the following condition:

$$y_m^c = \begin{cases} 1 & v_m^c \geq a \\ 0 & v_m^c < a \end{cases}, \tag{1}$$

where $a$ is the top-$K$ confidence value of $\mathbf{v}_m$.

**Smooth labeling.** Zheng *et al.* [12] propose a smooth labeling method for outliers which assigns an averaged label for each unlabeled data. For a generated image $\mathbf{g}_m$, the label is denoted as $\mathbf{y}_m = [y_m^1, y_m^2, \cdots, y_m^c, \cdots, y_m^C]^T$, where $C$ is the dimension of multi-label, $c = 1, 2, \cdots, C$. Each dimension of smooth multi-label $\mathbf{y}_m$ is set to a fixed value, *i.e.*, $y_m^c = \frac{1}{C}$.

**Soft pseudo labeling (SPL).** As previously described, pseudo labeling provides a definite category for unlabeled data. In this paper, we propose a flexible label mechanism called soft pseudo labeling (SPL), which assigns the virtual and appropriate multi-label for each generated unlabeled data. The prominent characteristic of the proposed label assignment method is that all label confidence values are considered rather than only over the pre-fixed threshold or the top-$K$ confidence values are concerned in traditional pseudo labeling [23]. Specifically, first, the initial confidence value $\mathbf{v}_m$ of an unlabeled GAN image $\mathbf{g}_m$ is obtained by passing the sample through a pre-trained model. Then, a virtual and appropriate value $\mathbf{y}_m$ is calculated by SoftMax function based on $\mathbf{v}_m$, *i.e.*, the $c$-th element of $\mathbf{y}_m$ is denoted as follows:

$$y_m^c = \frac{e^{v_m^c}}{\sum\limits_{c=1}^{C} e^{v_m^c}}, \tag{2}$$

where $C$ is the dimension of $\mathbf{y}_m$. Finally, the $\mathbf{y}_m$ is regarded as the soft pseudo label of the unlabeled GAN image $\mathbf{g}_m$. SoftMax activation function could establish a virtual relationship within independent labels and optimize label prediction. Therefore, each generated image is set a virtual label. Moreover, there is a limit state that the generated images have no related significant object, *i.e.*, the confidence values are very low or close to 0. In this situation, the categories of the generated image will be formulated as a equal probability distribution by above SoftMax function.

### 2.4. Training and Testing Processes of SPL

The proposed SPL contains the training process and testing process, respectively.

During training process, original training set $\mathbf{D_o} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{N}$ combined with additional generated set $\mathbf{D_g} = \{\mathbf{g}_m, \mathbf{y}_m\}_{m=1}^{M}$ are fed into the network as input. Following the reference [24] proposed by Zhu *et al.*, the architecture of the CNN model takes ResNet-101 [9] as backbone and replaces the block structure with the one proposed in [25]. Besides, we modify the last fully-connected layer to have $C$ neurons to predict the $C$ categories. $C$ is the total number of the classes in the original training set. The cross-entropy loss function $\mathcal{L}$ is used for guiding the learning process of the CNN model $\mathbb{M}$, defined as below:

$$\mathcal{L}(\mathbf{y}_i, \hat{\mathbf{y}}_i) = -\frac{1}{C}(\sum_{c=1}^{C} \log{(\hat{\mathbf{y}}_i)} \cdot \mathbf{y}_i), \tag{3}$$

where $\mathbf{y}_i = [y_i^1, y_i^2, \cdots, y_i^C]^T$ and $\hat{\mathbf{y}}_i = [\hat{y}_i^1, \hat{y}_i^2, \cdots, \hat{y}_i^C]^T$ denote the ground-truth label and estimated label, respectively.

During testing process, the final predicted label $\hat{\mathbf{y}}_j$ of testing image $I_j$ is obtained by passing through the trained model $\mathbb{M}$ with sigmoid activation function and a threshold.

## 3. EXPERIMENT

### 3.1. Datasets and Evaluation Protocol

This paper evaluates the multi-label image classification performance on three public widely-used datasets, *i.e.*, WIDER-Attribute [20], NUS-WIDE [21] and MS-COCO [22].

**WIDER-Attribute** [20] is a multi-label image classification dataset, containing $28,340$ images for training and $29,177$ images for testing. For each image, a person is annotated with 1 to 14 attributes.

**NUS-WIDE** [21] is a large-scale multi-label classification dataset, containing totally $269,648$ images collected from Flickr. The dataset is manually labeled with 81 concepts, with 2.4 concept labels per image on average. Following the settings in SRN [24], we set $119,986$ images for training, $107,859$ images for testing and $19,200$ images for validation.

**MS-COCO** [22] is primarily built for object detection in the context of scene understanding. The training set consists of $82,783$ images, which includes common objects in the scenes. The objects are classified into 80 categories, with 2.9 object labels per image on average. Following the settings in SRN [24], we set $82,783$ image for training, $40,504$ images for testing and $9,600$ images for validation.

**Evaluation Protocol.** In this paper, we adopt the mean Average Precision (denoted as **mAP**), macro/micro F1-measure (denoted as **F1-C/F1-O**), macro/micro precision (denoted as **P-C/P-O**) and macro/micro recall (denoted as **R-C/R-O**) for evaluation as same as [24].

### 3.2. Experimental Details and Setups

The DCGAN package[1] and WGAN-GP package[2] are employed to train the GAN model using the samples in the original training set without any pre-processing. Adam optimization [26] algorithm is utilized with the parameter $\beta_1 = 0.5$. All the images are randomly flipped before training. Training is done after $1,000$ epochs. Caffe [27] package is adopted to train the CNN model. During training process, we modify the last fully-connected layer to have 14, 81 and 80 neurons for WIDER-Attribute, NUS-WIDE and MS-COCO, respectively. All the images are resized to $256 \times 256$ before being randomly cropped into $224 \times 224$ with random horizontal flipping. The batch-size is 24. During testing process, we get a $C$-dim label prediction to evaluate the performance of multi-label image classification by calculating the coincidence degree between the predicted label and ground-truth.

In this paper, the baseline is that only the images in the ordinary training set are used as input to network during training process, and follows the architecture of the CNN model (proposed by Zhu *et al.* [24]) which takes ResNet-101 [9] as backbone and replaces the block structure with the one proposed in [25].

### 3.3. Experiment Results

The experimental results are listed in Table 1, 2 and 3.

**a) Comparison with the baseline.** We can observe that the proposed SPL outperforms the baseline under adding the same number of unlabeled samples to the original training set. Specifically, when the unlabeled samples are generated by DCGAN, mAP is improved $+0.5\%$ (from $83.7\%$ to $84.2\%$), $+1.3\%$ (from $55.7\%$ to $57.0\%$), $+1.5\%$ (from $71.9\%$ to $73.4\%$) on WIDER-Attribute, NUS-WIDE, MS-COCO, respectively. When the unlabeled samples are generated by WGAN-GP, mAP is improved $+0.6\%$ (from $83.7\%$ to $84.3\%$, $+1.0\%$ (from $55.7\%$ to $56.7\%$), $+1.6\%$ (from $71.9\%$ to $73.5\%$), on above three datasets, respectively.

**Table 1**. Quantitative results (%) comparisons on the WIDER-Attribute dataset.

| Methods | All | | |
|---|---|---|---|
| | mAP | F1-C | F1-O |
| DHC [20] | 81.3 | - | - |
| Baseline | 83.7 | 72.9 | 79.5 |
| Random labeling (DCGAN) | 82.0 | 71.7 | 78.5 |
| Random labeling (WGAN-GP) | 82.1 | 71.9 | 78.6 |
| Pseudo labeling (DCGAN) | 82.9 | 72.3 | 79.0 |
| Pseudo labeling (WGAN-GP) | 75.0 | 63.8 | 75.0 |
| Smooth labeling (DCGAN) | 84.2 | 73.5 | 79.9 |
| Smooth labeling (WGAN-GP) | 84.1 | 73.5 | 79.7 |
| Our SPL (DCGAN) | 84.2 | 73.8 | 80.0 |
| Our SPL (WGAN-GP) | **84.3** | **74.1** | **80.1** |

**b) Comparison with other label assignment methods.** To evaluate the effectiveness of SPL, some other label assignment methods are conducted for comparison, *i.e.*, random labeling, pseudo labeling and smooth labeling. Three observations are summarized. *i)* The random labeling is not suitable for multi-label image classification task. With random label assignment, the performance is even slightly inferior to baseline. The reason may be that a large number of random labeling samples bring data pollution to the original training set. *ii)* The experimental result of pseudo labeling is also not satisfactory, where only the top-3 highest confidence labels are assigned to the unlabeled samples during training process. In this way, the correspondingly discriminative information of some low confidence values is ignored, producing a compromised performance. *iii)* Both the smooth labeling and SPL provide a significant improvement over the baseline, demonstrating the effectiveness in multi-label image classification task. However, previous work [12] of smooth labeling is not designed for multi-label image classification task. We make an attempt to introduce smooth label assignment for improving the performance of multi-label image classification. Smooth labeling method assigns average labels of the pre-defined categories for unlabeled samples, ignoring the responses of the different salient objects. While SPL formulates all the pre-defined categories with various weights and provides a comprehensive assignment for unlabeled samples. Specifically, the performance of SPL is slightly superior to smooth label assignment on NUS-WIDE. The reason may be that the smooth labeling does not reflect the category saliency well. On MS-COCO, SPL and smooth labeling produce a similar performance. The reason may be that the categories

[1] https://github.com/carpedm20/DCGAN-tensorflow/

[2] https://github.com/LynnHo/DCGAN-LSGAN-WGAN-WGAN-GP-Tensorflow/

**Table 2**. Quantitative results (%) comparisons on the NUS-WIDE dataset.

| Methods | All | | | | | | | Top-3 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mAP | F1-C | P-C | R-C | F1-O | P-O | R-O | F1-C | P-C | R-C | F1-O | P-O | R-O |
| WARP [28] | - | - | - | - | - | - | - | 33.5 | 31.7 | 35.6 | 53.9 | 48.6 | 60.5 |
| CNN-RNN [11] | - | - | - | - | - | - | - | 34.7 | 40.5 | 30.4 | 55.2 | 49.9 | 61.7 |
| Baseline | 55.7 | 51.2 | **65.8** | 46.8 | 71.8 | **76.1** | 67.9 | 43.2 | 45.3 | 52.1 | 61.3 | 55.4 | 68.6 |
| Random labeling (DCGAN) | 51.9 | 47.9 | 60.8 | 45.2 | 70.7 | 74.4 | 67.2 | 40.8 | 43.2 | 50.1 | 60.3 | 54.5 | 67.5 |
| Random labeling (WGAN-GP) | 53.4 | 48.5 | 63.3 | 44.6 | 71.2 | 75.4 | 67.5 | 42.1 | **46.0** | 50.2 | 60.9 | 55.0 | 68.1 |
| Pseudo labeling (DCGAN) | 53.0 | 48.1 | 63.6 | 43.7 | 71.0 | 75.4 | 67.2 | 41.7 | 44.9 | 49.4 | 60.6 | 54.8 | 67.9 |
| Pseudo labeling (WGAN-GP) | 52.5 | 47.8 | 62.6 | 43.8 | 70.9 | 75.1 | 67.1 | 40.6 | 42.9 | 49.5 | 60.5 | 54.7 | 67.7 |
| Smooth labeling (DCGAN) | 55.7 | 52.4 | 63.0 | 49.7 | 71.7 | 74.7 | 69.0 | 43.9 | 45.3 | 53.6 | 61.2 | 55.3 | 68.4 |
| Smooth labeling (WGAN-GP) | 55.8 | 52.4 | 62.0 | 49.3 | 71.7 | 75.2 | 68.5 | 44.1 | 45.2 | 53.9 | 61.3 | 55.4 | 68.5 |
| Our SPL (DCGAN) | **57.0** | **53.4** | 63.6 | **50.9** | **72.2** | 74.4 | **70.1** | **44.5** | 45.6 | **54.3** | **61.4** | **55.5** | **68.7** |
| Our SPL (WGAN-GP) | 56.7 | 53.0 | 63.9 | 50.2 | 72.1 | 74.8 | 69.6 | 44.3 | 45.3 | 53.9 | **61.4** | **55.5** | **68.7** |

**Table 3**. Quantitative results (%) comparisons on the COCO dataset.

| Methods | All | | | | | | | Top-3 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mAP | F1-C | P-C | R-C | F1-O | P-O | R-O | F1-C | P-C | R-C | F1-O | P-O | R-O |
| WARP [28] | - | - | - | - | - | - | - | 55.7 | 59.3 | 52.5 | 60.7 | 59.8 | 61.4 |
| CNN-RNN [11] | - | - | - | - | - | - | - | 60.4 | 66.0 | 55.6 | 67.8 | 69.2 | **66.4** |
| Baseline | 71.9 | 65.8 | 79.8 | 58.6 | 71.8 | 81.9 | 63.8 | 62.2 | 82.7 | 53.4 | 69.2 | 85.9 | 58.0 |
| Random labeling (DCGAN) | 70.8 | 65.2 | 77.4 | 59.1 | 71.1 | 80.2 | 63.9 | 61.7 | 80.5 | 53.7 | 68.7 | 84.3 | 58.0 |
| Random labeling (WGAN-GP) | 70.6 | 65.0 | 77.2 | 59.0 | 71.0 | 80.3 | 63.7 | 61.6 | 80.2 | 53.7 | 68.7 | 84.5 | 57.8 |
| Pseudo labeling (DCGAN) | 70.4 | 64.7 | 77.5 | 59.3 | 71.0 | 80.5 | 63.4 | 61.4 | 80.8 | 53.4 | 68.5 | 84.4 | 57.5 |
| Pseudo labeling (WGAN-GP) | 70.2 | 64.5 | 78.1 | 57.6 | 70.7 | 80.9 | 62.8 | 61.1 | 81.0 | 52.6 | 68.3 | 84.8 | 57.1 |
| Smooth labeling (DCGAN) | 73.2 | 67.3 | 79.3 | 61.0 | 72.6 | 81.7 | 65.4 | 63.8 | 82.4 | 55.5 | 70.1 | 85.8 | 59.3 |
| Smooth labeling (WGAN-GP) | 72.8 | 67.0 | 77.8 | 60.6 | 72.4 | 81.5 | 65.2 | 63.5 | 82.3 | 55.1 | 69.9 | 85.7 | 59.0 |
| Our SPL (DCGAN) | 73.4 | 67.6 | 77.8 | **61.5** | 72.8 | 81.3 | **66.0** | 64.1 | 82.5 | **55.9** | 70.3 | 85.7 | 59.7 |
| Our SPL (WGAN-GP) | **73.5** | **67.8** | **79.9** | 61.3 | **72.9** | **82.1** | 65.6 | **64.4** | **83.3** | 55.8 | **70.4** | **86.4** | 59.4 |

of MS-COCO are more balanced.

c) **Comparison with some existing multi-label image classification methods.** We compare three existing multi-label image classification methods (*i.e.*, DHC [20], WARP [28] and CNN-RNN [11]). The proposed SPL brings decent improvement of benchmark on mAP, F1 and Precision in most cases. Experimental results demonstrate the effectiveness of the proposed SPL method.

## 4. CONCLUSION

In this paper, we propose a soft pseudo labeling (SPL) method to make full use of unlabeled data generated by GAN for multi-label image classification. Specifically, the unlabeled data is generated by two GAN models (*i.e.*, DCGAN [14] and WGAN-GP [17]). The virtual multiple labels of the generated unlabeled samples are assigned based on an initial confidence value by SoftMax function. In this way, the generalization ability of the learned CNN model is significantly improved. Meanwhile, some other hypothetical label assignment proposals are also discussed, providing candidate solutions of utilizing unlabeled data. On three public widely-used multi-label image classification datasets (*i.e.*, WIDER-Attribute [20], NUS-WIDE [21] and MS-COCO [22]), our method demonstrates consistent improvement over the corresponding baseline and produces a competitive performance compared with some existing multi-label image classification methods.

## 5. REFERENCES

[1] Xiaoya Wei, Ziwei Yu, Changqing Zhang, and Qinghua Hu, "Ensemble of label specific features for multi-label classification," in *Proc. ICME*, 2018, pp. 1–6.

[2] Tanfang Chen, Shangfei Wang, and Shiyu Chen, "Deep multimodal network for multi-label classification," in *Proc. ICME*, 2017, pp. 955–960.

[3] Jian Wu, Shiquan Zhao, Victor S Sheng, Pengpeng Zhao, and Zhiming Cui, "Multi-label active learning for image classification with asymmetrical conditional dependence," in *Proc. ICME*, 2016, pp. 1–6.

[4] Jing Shao, Kai Kang, Chen Change Loy, and Xiaogang Wang, "Deeply learned attributes for crowded scene understanding," in *Proc. CVPR*, 2015, pp. 4657–4666.

[5] Foteini Markatopoulou, Vasileios Mezaris, and Ioannis Patras, "Implicit and explicit concept relations in deep neural networks for multi-label video/image annotation," *IEEE Trans. on Circuits and Systems for Video Technology*, 2018.

[6] Haofu Liao, Yuncheng Li, and Jiebo Luo, "Skin disease classification versus skin lesion characterization: Achieving robust diagnosis using multi-label deep neural networks," in *Proc. ICPR*, 2016, pp. 355–360.

[7] Yingying Zhu, Xiaofeng Zhu, Minjeong Kim, Jin Yan, Daniel Kaufer, and Guorong Wu, "Dynamic hypergraph inference framework for computer assisted diagnosis of neurodegenerative diseases," *IEEE Trans. on Medical Imaging*, vol. 38, no. 2, pp. 608–616, 2019.

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, 2016, pp. 770–778.

[10] Yunchao Wei, Wei Xia, Min Lin, Junshi Huang, Bingbing Ni, Jian Dong, Yao Zhao, and Shuicheng Yan, "Hcp: A flexible cnn framework for multi-label image classification," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 38, no. 9, pp. 1901–1907, 2016.

[11] Jiang Wang, Yi Yang, Junhua Mao, Zhiheng Huang, Chang Huang, and Wei Xu, "Cnn-rnn: A unified framework for multi-label image classification," in *Proc. CVPR*, 2016, pp. 2285–2294.

[12] Zhedong Zheng, Liang Zheng, and Yi Yang, "Unlabeled samples generated by gan improve the person reidentification baseline in vitro," in *Proc. ICCV*, 2017, pp. 3774–3782.

[13] Zhun Zhong, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang, "Camera style adaptation for person reidentification," in *Proc. CVPR*, 2018, pp. 5157–5166.

[14] Alec Radford, Luke Metz, and Soumith Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. ICLR*, 2016.

[15] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. ICCV*, 2017, pp. 2242–2251.

[16] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *Proc. NIPS*, 2014, pp. 2672–2680.

[17] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville, "Improved training of wasserstein gans," in *Proc. NIPS*, 2017, p-p. 5767–5777.

[18] Wu Liu, Xinchen Liu, Huadong Ma, and Peng Cheng, "Beyond human-level license plate super-resolution with progressive vehicle search and domain priori gan," in *Proc. ACM MM*, 2017, pp. 1618–1626.

[19] Shuang Ma, Jianlong Fu, Chang Wen Chen, and Tao Mei, "Da-gan: Instance-level image translation by deep attention generative adversarial networks," in *Proc. CVPR*, 2018, pp. 5657–5666.

[20] Yining Li, Chen Huang, Chen Change Loy, and Xiaoou Tang, "Human attribute recognition by deep hierarchical contexts," in *Proc. ECCV*, 2016, pp. 684–700.

[21] Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yantao Zheng, "Nus-wide: a real-world web image database from national university of singapore," in *Proc. ACM ICIVR*, 2009.

[22] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, "Microsoft coco: Common objects in context," in *Proc. ECCV*, 2014, pp. 740–755.

[23] Dong-Hyun Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. ICML Workshop*, 2013.

[24] Feng Zhu, Hongsheng Li, Wanli Ouyang, Nenghai Yu, and Xiaogang Wang, "Learning spatial regularization with image-level supervisions for multi-label image classification," in *Proc. CVPR*, 2017, pp. 2027–2036.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Identity mappings in deep residual networks," in *Proc. ECCV*, 2016, pp. 630–645.

[26] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2014.

[27] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. ACM MM*, 2014, pp. 675–678.

[28] Yunchao Gong, Yangqing Jia, Thomas Leung, Alexander Toshev, and Sergey Ioffe, "Deep convolutional ranking for multilabel image annotation," in *Proc. ICLR*, 2014.