Unsupervised Deep Feature Transfer for Low Resolution Image Classification

Zhang, Z.; Wu, Y.; Wang, G.
TR2019-127   October 31, 2019

Abstract

In this paper, we propose a simple while effective unsupervised deep feature transfer algorithm for low resolution image classification. No fine-tuning on convenet filters is required in our method. We use pre-trained convenet to extract features for both high and low resolution images, and then feed them into a two-layer feature transfer network for knowledge transfer. A SVM classifier is learned directly using these transferred low resolution features. Our network can be embedded into the state-of-the-art deep neural networks as a plug-in feature enhancement module. It preserves data structures in feature space for high resolution images, and transfers the distinguishing features from a wellstructured source domain (high resolution features space) to a not well-organized target domain (low resolution features space). Extensive experiments on VOC2007 test set show that the proposed method achieves significant improvements over the baseline of using feature extraction.

IEEE International Conference on Computer Vision Workshops (ICCV)
Unsupervised Deep Feature Transfer for Low Resolution Image Classification

Yuanwei Wu∗, Guanghui Wang
EECS, The University of Kansas
1450 Jayhawk Blvd., Lawrence, KS 66045
{y262w558, ghwang}@ku.edu

Ziming Zhang†
Mitsubishi Electric Research Laboratories
201 Broadway, Cambridge, MA 02139-1955
zzhang@merl.com

Abstract
In this paper, we propose a simple while effective unsupervised deep feature transfer algorithm for low resolution image classification. No fine-tuning on convenet filters is required in our method. We use pre-trained convenet to extract features for both high and low resolution images, and then feed them into a two-layer feature transfer network for knowledge transfer. A SVM classifier is learned directly using these transferred low resolution features. Our network can be embedded into the state-of-the-art deep neural networks as a plug-in feature enhancement module. It preserves data structures in feature space for high resolution images, and transfers the distinguishing features from a well-structured source domain (high resolution features space) to a not well-organized target domain (low resolution features space). Extensive experiments on VOC2007 test set show that the proposed method achieves significant improvements over the baseline of using feature extraction.

1. Introduction

Recently, deep neural networks (DNNs) have demonstrated impressive results in image classification [15, 12], object detection [7, 22, 18], and instance segmentation [10]. The success of DNNs has become possible mostly due to a large amount of available labeled data [5], as well as advances in better learning algorithms [8, 28] and computing resources. All those work typically assume that images are sufficiently high resolution (e.g. 224 × 224).

People in the research communities have overcome the limitation of DNNs, which requires large amount of data to train, by introducing transfer learning techniques. A common way to make use of transfer learning in the context of DNNs is to start from a pre-trained model in a similar task or domain, and then finetune the parameters to the new task. For example, the pre-trained model on ImageNet for classification can be finetuned for object detection on Pascal VOC [7, 22].

In this paper, we focus on low resolution (e.g. 32 × 32 or less) image classification as for privacy purpose, it is common to use low resolution images in real-world applications, such as face recognition in surveillance videos [29]. Without additional information, learning from low resolution images always reduces to an ill-posed optimization problem, and achieves a much degraded performance [21].

As shown in Fig. 1, the deep feature of high resolution images extracted from pre-trained convenet has already learned discriminative per-class feature representation. Therefore, it is able to be well separated in the tSNE visualization. However, the extracted feature of low resolution images are mixed together. A possible solution is to exploit the transfer learning, leveraging the discriminative feature representation from high resolution images to low resolution images.

In this paper, we propose a simple while effective unsupervised deep feature transfer approach that boosts classification performance in low resolution images. We assume that we have access to high resolution labeled images during training, but at test we only have low resolution images. Most existing datasets are high resolution. Moreover, it is much easier to
label subcategories in high resolution images. Therefore, we believe it is a reasonable assumption. We aim to transfer knowledge from such high resolution images to real world scenarios that only have low resolution images. The basic intuition behind our approach is to utilize high quality discriminative representations in the training domain to guide feature learning for the target low resolution domain.

The contributions of our work have three-fold.

- No fine-tuning on convenet filters is required in our method. We use pre-trained convenet to extract features for both high resolution and low resolution images, and then feed them into a two-layer feature transfer network for knowledge transfer. A SVM classifier is learned directly using these transferred low resolution features. Our network can be embedded into the state-of-the-art DNNs as a plug-in feature enhancement module.

- It preserves data structures in feature space for high resolution images, by transferring the discriminative features from a well-structured source domain (high resolution features space) to a not well-organized target domain (low resolution features space).

- Our performance is better than that of baseline using feature extraction approach for low resolution image classification task.

2. Related Work

Our method is closely related to unsupervised learning of features and transfer learning.

Unsupervised learning of features: Clustering has been widely used for image classification [2, 26, 13]. Ji et al. [13] propose invariant information clustering relying on statistical learning by optimising mutual information between related pairs for unsupervised image classification and segmentation. Caron et al. [2] present a clustering method that jointly learns the parameters of a neural network and the cluster assignments of the resulting features. Yang et al. [26] propose an approach to jointly learn deep representations and image clusters by combining agglomerative clustering with CNNs and formulate them as a recurrent process.

Transfer learning: It is commonly used in the scenario where the training and testing data distributions are different. Saenko et al. [24] learn a regularized non-linear transformation in the context of object recognition to minimize the effect of domain-induced changes in the feature distribution. Chen et al. [4] transfer knowledge stored in one previous network into each new deeper or wider network to accelerate the training of a significantly larger neural network. Yosinski et al. [27] experimentally study the transferability of hierarchical features in deep neural networks. Azizpour et al. [1] investigate the factors of transferability of a generic deep convolutional networks such as the network architecture, distribution of the training data, etc. Tzeng et al. [25] learn a CNN architecture to optimize domain invariance and transfer information between tasks. Long et al. [17] propose a deep adaptation network architecture to match the mean embeddings of different domain distributions in a reproducing kernel Hilbert space. Guo et al. [9] propose an adaptive fine-tuning approach to find the optimal fine-tuning strategy per instance for the target data. Readers can refer to [20] and the references therein for details about transfer learning.

3. Proposed Approach

After a short introduction to the supervised learning of deep neural networks, we describe our unsupervised deep feature transfer approach.
3.1. Preliminary

With the recent success of deep learning in computer vision, the deep convnets have become a popular choice for representation learning, to map raw images to an embedding vector space of fixed dimensionality. In the context of supervised learning, they could achieve better performance than human beings on standard classification benchmarks \[11, 15\] when trained with large amount of labelled data.

Let \( f_\theta \) denote the convenet mapping function, where \( \theta \) is the corresponding learnable parameters. We refer to the vector obtained by applying this mapping to an image as feature or features. Given a training set \( X = \{x_1, \cdots, x_N\} \) of \( N \) images, and the corresponding ground truth labels \( Y = \{y_1, \cdots, y_N\} \), we want to find an optimal parameter \( \theta^* \) such that the mapping \( f_\theta \) predicts good general features. Each image \( x_i \) associates with a class label \( y_i \in \{0, 1\}^k \). Let \( g_\omega \) denote a classifier with parameter \( \omega \). The classifier would predict the labels on top of the features \( f_\theta(x_i) \). The parameter \( \theta \) of the mapping function and the parameter \( \omega \) of the classifier are then learned jointly by optimizing the following objective function:

\[
\min_{\theta, \omega} \frac{1}{N} \sum_{i=1}^{N} L(g_\omega(f_\theta(x_i), y_i)), \tag{1}
\]

where \( L \) is the multinominal logistic loss for measuring the difference between the predicted labels and ground-truth labels given training data samples.

3.2. Unsupervised Deep Feature Transfer

The idea of this work is to boost the feature learning for low resolution images by exploiting the capability of unsupervised deep feature transfer from the discriminative high resolution feature. The overview of proposed approach is shown in Fig. 2. It consists of three modules: feature extraction, unsupervised deep feature transfer, and classification, discussed below.

**Feature extraction.** We observe that the deep features extracted from convenet could generate well separated clusters as shown in Fig. 1. Therefore, we introduce the transfer learning to boost the low resolution features learning via the supervision from high resolution features. Then, we extract the features (N-Dimensional) of both high and low resolution images from a pre-trained deep convenet. More details are described in Sec. 4.2.

**Unsupervised deep feature transfer.** We propose a feature transfer network to boost the low resolution features learning. However, in our assumption, the ground truth labels for low resolution images are absent. Therefore, we need to make use of the information from high resolution features. In order to do this, we propose to cluster the high resolution features and use the subsequent cluster assignments as “pseudo-label” to guide the learning of feature transfer network with low resolution features as input. Without loss of generality, we use a standard clustering algorithm, k-means. The k-means takes a high resolution feature as input, in our case the feature \( f_\theta(x_i) \) extracted from the convenet, and clusters them into \( k \) distinct groups based on a geometric criterion. Then, the pseudo-label of low resolution features are assigned by finding its nearest neighbor to the \( k \) centroids of high resolution features. Finally, the parameter of the feature transfer network is updated by optimizing Eq. (1) with mini-batch stochastic gradient descent.

**Classification.** The final step is to train a commonly used classifier such as Support Vector Machine (SVM) using the transferred low resolution features. In testing, given only the low resolution images, first, our algorithm extracts the features. Then feeds them to the learned feature transfer network to obtain the transferred low resolution features. Finally, we run SVM to get the classification results directly.

4. Experiments

4.1. Dataset

We conduct the low resolution classification on the PASCAL VOC2007 dataset \[6\] with 20 object classes. There are 5,000 images in VOC2007 trainval set and 4,952 images in VOC2007 test set. However, the images in the dataset are high resolution images only. We follow \[16\] to generate the low resolution images. In this work, we generate high resolution images by resizing the original images to \( 224 \times 224 \) using bicubic interpolation. We generate the low resolution images by down-sampling the original to \( 32 \times 32 \), and then up-sampling to \( 224 \times 224 \).

4.2. Implementation Details

We conduct our experiment using Caffe \[14\]. We use the resnet-101 \[12\] pre-trained on ILSVRC2012 \[23\] as the backbone convenet to extract the features from high and low resolution images. We extract the features from the pool5 layer, which gives a feature vector with dimension of \( N = 2048 \).

The feature transfer network is a two-layer fully connected network. We conduct grid search to find the optimal design for the network architecture, see Sec. 4.3. It is initialized using MSRA \[14\] initialization. We train the feature transfer network using stochastic gradient descent with weight decay 0.0005, momentum 0.9, batch size 1,000, epoch 1,000, total iteration 31,561. The initial learning rate is 0.01, and is decreased by 10 after every 15,000 iterations.

4.3. Feature Transfer Network

The feature transfer network is shallow, with two fully connected layers. Let \( N_1 \) and \( N_2 \) denote the neurons of the

\[1\] We download the Caffe Model from https://github.com/BVLC/caffe/wiki/Model-Zoo
first and second fully connected layers, respectively. We conduct grid search to find the optimal combination for $N_1$ and $N_2$, as shown in Table 2. The number $N_2$ is determined by the number of clusters $k$ for the pseudo labels in k-means.

As we can see, when the neurons of $N_2$ is fixed, the mAP increases as the neurons of $N_1$ increases. This is because the capacity of the two-layers feature transfer network increases as the neurons increases in $N_1$. However, given a fixed number of neurons of $N_1$, the value of mAP would increase first, and then decrease when the value of neurons in $N_2$ is larger enough, maybe 200 is a threshold value in our two-layer network as shown in the table. We observe that the hyperparameters with $N_2 = 100$ and $N_1 = 4096$ for the neurons give us the best performance. We use the same values in our experiment.

### 4.4. Low Resolution Image Classification

We evaluate the performance of image classification in the context of binary classification task on the VOC2007 test set using SVM [3] classifier in matlab. We have compared our algorithm with two baselines: Baseline-HR and Baseline-LR, discussed below. Baseline-HR is to use the extracted high resolution features (2048-D) of VOC2007 trainval set to train the SVM and report the classification performance on VOC2007 test set. It is similar for Baseline-LR, but with the extracted low resolution features (2048-D). Our method transfers the low resolution feature from 2048-D to 100-D. Therefore, we train the SVM using the 100-D features for each class. We show the comparison in Table 1.

The Baseline-HR is the upper bound of our method, and Baseline-LR is the lower bound. As we can see from the Table 1, the proposed unsupervised deep feature transfer is able to boost the low resolution image classification by about 2%. Except for the classes of “bottle” and “sheep”, our method outperforms the Baseline-LR. As shown in Fig. 3, we find the transferred low resolution features are separated much better than the extracted low resolution features. Those indicate that the proposed unsupervised deep feature transfer algorithm does help transfer more discriminative representations from high resolution features. Therefore, it boost on low resolution images classification task. The feature transfer network could also be embedded into the state-of-the-art deep neural networks as an plug-in module to enhance the learned features.

### 5. Conclusion

In this paper, we propose a unsupervised deep feature transfer algorithm for low resolution image classification. The proposed two-layer feature transfer network is able to
boost the classification by 2% on mAP. It can be embedded into the state-of-the-art deep neural networks as a plug-in feature enhancement module. While our current experiments focus on generic classification, we expect our feature enhancement module to be very useful in detection, retrieval, and category discovery settings as well in the future.

References


