Rethinking ImageNet Pretraining in Domain Adaptation

Junzhi Ning

February 19, 2022

Abstract

1 Introduction

Unsupervised domain adaptation(UDA) is an study area of transfer learning dedicated to construct the model that is able to achieve the tasks by adapting a target domain from the source domain. Today, two main approaches of UDA are Domain-invariant feature learning and domain mapping. Domain-invariant feature learning[14] aims to align the source and target domain by generating a feature representation common to both domains. This approach typically defines a divergence used to measure the distance between two domains at the level of feature representation and then attempts to find the most appropriate one that can minimize that divergence. Alternatively, the domain mapping[12]in UDA designs to find the best mapping directly from one domain to another so that the input feature can be mapped into the domain that has known labels to train a classifier.

In recent years, upon the burgeoning development of neural network-based model and the increasing success of Generative Adversarial model[4], the adversarial-based methods have also been introduced into the field of UDA[1, 9, 3]. However, the NN-based models often under-perform in UDA tasks when the training data from the source and target domains is insufficient. Current methods mostly address this problem by using the pre-trained networks from the larger training dataset as the initialization weights and it serves as a starting point of many existing models to tackle problems in applications. Specifically, in computer vision tasks, pre-trained networks with various architectures from ImageNet[2] are usually preferred. Nonetheless, due to the large number of classes in ImageNet, the overall effect brought by pre-trained networks on the performance of the UDA tasks remains unclear and it is possible for the pre-trained networks to play an non-trivial role in reducing the domains gap, thereby overestimating the intrinsic functionality of current domain adaptation structure.

A recent study [7] shows that removing a portion of classes in the pretrained model from the ImageNet dataset up to 20 %, can still achieve comparable or even better performance in transfer learning. However, the experiment from this paper only studies the effect of the number of classes from ImageNet dataset used in the pretrained model, the choice of removing or ignoring certain classes when using the pretrained model weights in specific UDA classification tasks like Office31 dataset has yet been explored.

Motivated by the above discussion, we will investigate how ImageNet pretraining affect the domain adaptation methods. Particularly, we will look into muting some chosen ImageNet class labels of same or similar types in the UDA benchmark datasets when obtaining the pre-trained model.

2 Related Work and Concepts

In this section, we reviews basic and important concepts throughout our study and experiments.

2.1 Alex-net

Alexnet which is a name of CNN architecture was proposed by Alex Krizhevsky in the paper[6] of 2012, it participated ImageNet Large Scale Visual Recognition Challenge in 2012 and achieved top-5 error rate of 15.3%.The depth of the alexnet model is key to its high performance. It contains 8 main layers with learn-able weights. The first five layers are Convolutional layers, for each of five Convolutional layers, it is followed by one RELU activation function and one layer of maxpooling, these five layers serve as the feature extractor. The last three layers are fully connected, and the output layer produces 1000-softmax float values in aim of achieving the property to simulate the probability distribution.

In our experiments, we modify the output layer of the Alexnet to adjust the number of masked class labels for imagenet pretraining and then apply it to be the feature extractor of UDA and SSL methods in fine-tuning phase.

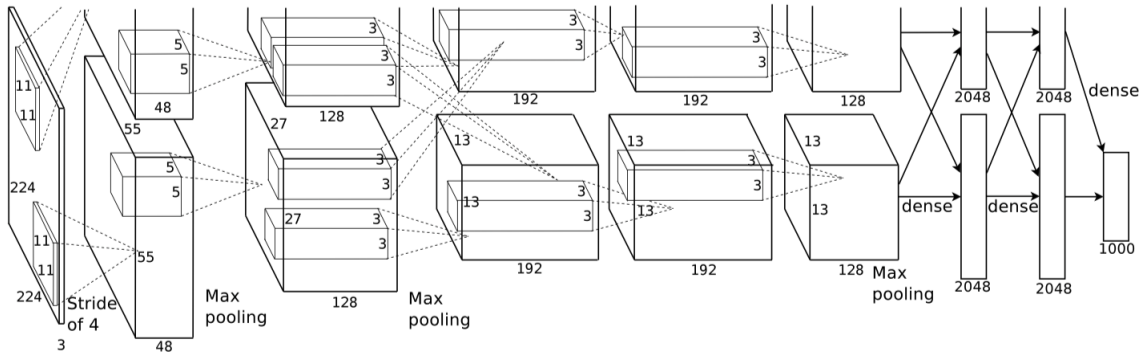

Figure 1: Graphical illustration of Alex-net [6]

2.2 Unsupervised Domain Adaptation (UDA)

In a classification task of UDA[10, 13], given n_s labeled samples as a set $D_s = \{x_j, y_j\}_{j=1}^{n_s}$ from source domain with probability distribution of $P_s(X, Y)$ and n_t unlabeled samples as a set $D_t = \{x_j\}_{j=1}^{n_t}$ from target domain with probability distribution of $P_t(X)$ which is the marginal probability distribution of X without considering the predefined label set Y and we assume $\forall x \in D_t$, each x has a unknown corresponding label $y' \in Y$, the main goal of UDA classification task is to build a classifier C that minimizes the target risk $\mathbb{E}_{(x,y')\sim P_t(X,Y)}|C(x) - y'|$.

2.2.1 Domain-adversarial Neural Network(DANN)

DANN was originally proposed in the paper [3] of 2016 by Yaroslav Ganin and his colleagues to tackle the problem in DA classification. It is based on the main ideas of adversarial-based network and Domain-invariant feature learning, it create a model composed of three parts, through an adversarial approach so that the feature representation generated by the feature extractor, a component of the DANN, can not be discriminated effectively by the domain classifier, then the label predictor employs the obtained feature representation to produce the corresponding class label for both source and target

domains. Figure 2

To be more specific, the DANN promote the feature representation that has following properties:

- discriminative for the classification task on the source domain.
- indiscriminative with respect to the divergence between the source and target domains.

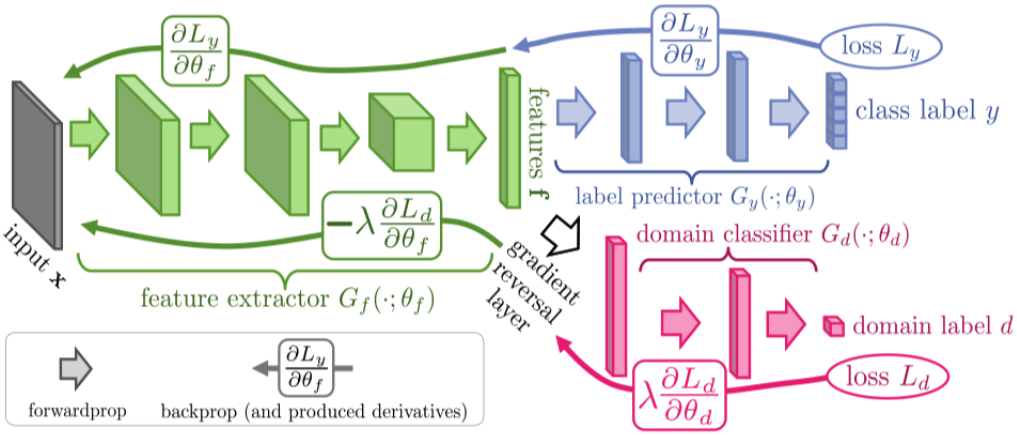

Figure 2: Graphical illustration of DANN [3]

To complement the study of the effect of pretrained network weight on UDA, we also briefly mention the notion of semi-supervised learning in order to see if it has similar scale of dependence on pretrained network in SSL setting, but it will not be main focus of our experiments.

A research[13] in 2021 suggests that the SSL methods outperform existing UDA methods on the UDA benchmark and therefore should be promoted as baselines in future. Since the research paper also employs the pretrained weights as initialization for training SSL methods, we are also interested in finding out the impact of feature representation of existing classes in pretrained networks on the SSL methods.

2.3 Semi-supervised Learning

In a classification of SSL, given n_l labeled samples as a set $D_l = \{x_i, y_i\}_{i=1}^{n_l}$ and n_u unlabeled samples as a set $D_u = \{x_i\}_{i=1}^{n_u}$ from probability distributions of $P(X)$ and $P(X,Y)$ respectively, typically assuming the size of D_u is less than the size of D_l, the main objective of SSL is to find the a mapping C such that it minimizes the risk $\mathbb{E}_{(X,Y)\sim P(X,Y)}|C(x) - y|$.

2.3.1 Pseudo-label method

The Pseudo-label method was introduced in the paper [8] of the Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks in 2013. It trains the network with labeled and unlabeled data simultaneously. During the training iterations, the unlabeled data will being gradually labeled with class that has highest probability and confidence. The measure of confidence and probability can be calculated through applying the Sigmoid Unit or softmax Unit to final output. In the original paper, it mentions that The Pseudo-label method in principle equivalent to Entropy Regularization

[5] and the resulting effect of minimizing the entropy for unlabeled data exhibits the favour to a low-separation between classes. In our study, we utilize this method to conduct the experiments for the aspect of SSL setting.

3 Experiments

3.1 Brief Summary of Workflow

For the first part of our experiment, we manually scrutinise through the label list of the pre-trained dataset of ILSVRC and attempt to filter out class labels that are similar or exactly matched with categories presented in the dataset of Office31. Then, by ignoring those chosen classes in ILSVRC face-blurred dataset, we adopt an Alexnet model to classify the ILSVRC face-blurred dataset. At the same time, the same number of class labels in the ILSVRC face-blurred dataset is randomly chosen to train another alexnet model in order to replicate the training environment of the pretrained model on the chosen masked label list, this serves as a control group to evaluate the difference.

For the second part of our experiment, we adopt two Alexnet models pretrained on the ILSVRC face-blurred dataset as the feature extractors after removing last few layers of the models. The feature extractors are then embedded into DANN and Pseudo-label methods respectively to perform the fine tuning in domain adaptation on three domains of Office31. We set most of hyper-parameters according to either following the original papers or available machine learning libraries, we will look into details later in this section.

3.2 Datasets

3.2.1 ImageNet (ILSVRC face-blurred)[2]

ImageNet is a dataset with a size of over 14 million labeled images, categorising into more than 20000 classes. The images from the ImageNet dataset were collected from online resources and they were manually labeled for uses in computer vision applications and research. Since 2010, in light of Pascal Visual Object Classes (VOC) challenge, the competition known as The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) takes place annually. A subset of images from the ImageNet database, consisting of more than 1.2 million training labeled images of 1000 category, roughly 50 thousand validation images and 150 thousand testing images are chosen as their competition datasets.

For the purpose of our experiments, we uses the latest version of ILSVRC ImageNet face-blurred dataset for training our pretrained models.

3.2.2 Office31

The office31 is a dataset consisting of 31 object classes in three domains. The 31 object classes are common tools and items in daily office setting. The Amazon domain contains a total of over 2800 images with a resolution, most of images are captured by amazon online merchants. The DSLR domain consists of around 500 images with a high resolution of 4288×2848, each class contains 5 images and each object are captured from different angles. For the Webcam domain, 795 low resolution images are shared across 31 categories with a significant noise and color. Due to availability of multiple domains in the dataset, the Office31 are widely adopted in field of transfer learning for comparison purposes and evaluation of model performance.

Figure 3: Examples of ImageNet images

For the purpose of our experiments, we use Office31 dataset to perform the benchmark evaluation for UDA classification.

(a) Amazon domain (b) Dslr domain (c) Webcam domain

Figure 4: Examples of images from Office31

3.3 Setups

Here we adopt the public Pytorch library [11] of Image-Net example as a starting point of our experiment. We create three seperate datasets from the ILSVRC Image-Net face-blurred dataset of Table 1.

Data-sets	No.classes	No.masked classes	No. Training images	No.validation images
Original	1000	0	1281066	49997
Masked	958	42	1226767	47897
Masked Random	958	42	1227158	47897

Table 1: Masked label list attached in Appendix 3.
Note that every class label has the same number of validation images.

To enhance the models' robustness and generality, we add image transformations to the training images before feeding to the models, these following transformations are Horizontally flipping an image with a given probability, resizing an image of 256×256, cropping an image at a random point and resizing it into a size of 224×224.

The implementation of Alexnet in our experiments does not follow the default version in pytorch library, but instead we use the modified version from the paper [9], this version of Alexnet is easier to converge and also covers the implementation of Local Response Normalization layer of which the default version in pytorch omits. Three alexnet models are trained on the three datasets as pretrained models1 using the following setting 2 in training.

No.Epochs	learning rate	Optimizer	Batch size	Weight decay	LR scheduler
100	0.01	SGD	256	0.0005	30 epochs rate decay of 0.1

Table 2: Models training hyper-parameters and setting.
Note that SGD stands for Stochastic Gradient Descent

3.4 Results

4 Discussions

5 Conclusions

References

[1] Konstantinos Bousmalis et al. "Unsupervised pixel-level domain adaptation with generative adversarial networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 3722–3731.

[2] Jia Deng et al. "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.

[3] Yaroslav Ganin et al. "Domain-adversarial training of neural networks". In: *The journal of machine learning research* 17.1 (2016), pp. 2096–2030.

[4] Ian Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems* 27 (2014).

[5] Yves Grandvalet and Yoshua Bengio. *Entropy Regularization.* 2006.

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems* 25 (2012).

[7] Michal Kucer and Diane Oyen. "Transfer learning with fewer ImageNet classes". In: (2021).

[8] Dong-Hyun Lee et al. "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks". In: *Workshop on challenges in representation learning, ICML*. Vol. 3. 2. 2013, p. 896.

[9] Mingsheng Long et al. "Conditional adversarial domain adaptation". In: *Advances in neural information processing systems* 31 (2018).

[10] Mingsheng Long et al. "Learning transferable features with deep adaptation networks". In: *International conference on machine learning*. PMLR. 2015, pp. 97–105.

[11] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035. URL: `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

[12] Ashish Shrivastava et al. "Learning from simulated and unsupervised images through adversarial training". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 2107–2116.

[13] Yabin Zhang et al. "Semi-supervised models are strong unsupervised domain adaptation learners". In: *arXiv preprint arXiv:2106.00417* (2021).

[14] Han Zhao et al. "On learning invariant representations for domain adaptation". In: *International Conference on Machine Learning*. PMLR. 2019, pp. 7523–7532.

Appendix

ImageNet class label	Office31 class label
cup	
comic book	
wine bottle	
water bottle	
typewriter keyboard	
tray	
table lamp	
spotlight	
screen	
rocking chair	
projector	
printer	
pop bottle	
pill bottle	
pay phone	
notebook	
mouse	
mountain bike	
moped	
monitor	
microphone	
medicine chest	
loudspeaker	
laptop	
lampshade	
hand-held computer	
fountain pen	
football helmet	
folding chair	
dial telephone	
desktop computer	
crash helmet	
computer keyboard	
coffee mug	
china cabinet	
cellular telephone	
bottlecap	
bookcase	
binder	
bicycle	
beer bottle	
barberchair	

Table 3: Masked label list for ImageNet.