Graph algorithm optimization techniques for high-throughput computers in weakly supervised scenarios

Liping Mao*
Nanchang Institute of Technology Nanchang 330044 China
*Corresponding author: maoliping_nclg@nut.edu.cn

Abstract. In this paper, a weakly supervised recognition model is used to provide approximate position constraints for object transformation in the source and target domains, and to guide the generator to generate images that differ less from the real ones. The cross-modal synthesis technique is incorporated to take advantage of the data differences between different modalities to constrain the mapping functions, thus ensuring that the learned mapping functions are always performed on the paired data. The full version of the proposed method achieves a 3.6% performance improvement compared to the benchmark network. Richly controlled experiments demonstrate the effectiveness of each improvement. It is found that the intrinsic correlation of query graphs provides complementary information for better performance of semantic segmentation with few samples, and the proposed network structure in this paper effectively exploits this information.

Keywords: weakly supervised scenes; high-throughput computers; image algorithms; optimization techniques

1. Introduction

With the recent rapid development of modern computer technology, machine learning has become an important approach in various neighbourhoods, where machine learning can be further classified into unsupervised learning, weakly supervised learning, and supervised learning based on the degree of labelling of the training data set. Weakly-supervised learning as compared to supervised learning, where supervised learning is when a pair of registered data as input for training and the model is fitted to the corresponding labelled results based on the data [1]. The second is semi-supervised learning, which requires that the labels in the same cluster in the sample space should be the same, and the labels of the samples close to each other should be the same as well. Semi-supervised learning tries to automatically process the unlabelled data and train the model without human intervention.

The goal of computer vision is to achieve image understanding. In computer vision, object classification and recognition occupy an important position [2]. The main problem to be solved for object classification is to determine whether there is a certain class of object in this image for a given image, i.e., the existence of the object; for object detection, the problem to be solved is to determine the type of object and calculate the location of the object in the image, i.e., marking out the rectangular box; object recognition, in contrast to object detection, only gives the approximate location of the existence of the object. The current problem in the field of object recognition and detection is that to
improve detection accuracy, the data samples needed for research often require a large number of manually labelled example-level rectangular boxes, but this manual labelling of object location rectangular boxes is not widely used in real-life production, on the one hand, the cost is too high, on the other hand, the labelling is more subjective and the error is relatively large. One of the trending research directions in the field of object recognition and detection is to use only weakly labelled data samples to obtain both class and location information of objects.

Image translation models cannot focus only on specific scene objects, for example, in the unsupervised case, if images are not paired or aligned, the network must additionally learn which parts of the scene to translate. To achieve a translation that works well for horses to zebras is to require the network to focus on each animal and to change only the animal part of the image. No additional supervised training is required for the entire network model.

2. Status of research

Smith et al. modelled the relationship between low-resolution images and high-resolution images by Markov networks, and then found the local maximum posterior probability in high-resolution images based on the Bayesian belief propagation algorithm, and reconstructed the images accordingly with good results [3]. Wagholikar et al. combined various learning-based super-resolution methods, assuming that the small blocks in the low-resolution image have similar geometry as those in the high-resolution image, and based on this a priori condition, the geometric features of the low-resolution image are used as the geometric features of the high-resolution image, and good reconstruction results are also achieved [4]. Based on the principle of compressed perception, Yang et al. regarded the low-resolution image as a down sampled version of the high-resolution image and assumed that the image blocks could be sparsely represented by a set of complete dictionaries, and then recovered the sub-blocks of the high-resolution image from these low-resolution image blocks, and this method could achieve good results when dealing with a certain class of images with the same properties [5].

Network design mainly includes the design of network structure and loss function [6].

Based on the regional suggestion network framework, highly robust image enhancement is firstly introduced to solve the processing problem of large images with high resolution. Meanwhile, rotated coordinates and multi-order feature fusion strategies are introduced to improve the advanced semantic feature mapping, which solves the dense target regions and reduces the redundant target detection regions. In this paper, a multi-branch random perturbation learning framework based on convolutional neural networks is proposed, while a metric function is introduced to enhance the learning ability among different branches. A delayed update strategy with random perturbation is proposed to avoid fast falling into local optimal video sequences. In the actual tracking process, perturbation factors are added to expand the solution space.

3. Design of high-throughput computer graph algorithms for weakly supervised scenes

3.1. Preparation of semi-rigid zinc coordination polymers

Objects may appear in different locations, at different scales, and from different viewpoints, and network models should focus on building images for these complex scenes to make them more adaptable to the needs of everyday production [7]. The purpose of transfer learning is to apply a model learned in a certain context to a different domain, migrating data or model structures from the source domain to the target domain to meet the task requirements of the target domain, as shown in Figure 1 for a schematic diagram of transfer learning. Theoretically, a fully-connected convolutional layer can replace layer if the convolutional kernel is of the same size as the input data. When replacing a fully connected layer with a convolutional layer, the number of convolutional layer parameters used remains the same, but the introduction of the convolutional layer allows the network to input images of arbitrary size on the one hand and makes the computation more efficient on the other.

If the sliding window is used, a sliding window of 224×224 should be chosen to slide through the input image in 32 steps, which will produce another 6×6 image, and after passing all 36 images
through the convolution and fully-connected layers again, the fraction of the original image at 36 positions can be obtained. If the last three fully-connected layers are replaced by convolutional layers, then without the sliding window, the $12 \times 12 \times 512$ feature map is obtained after 5 convolutional and pooling layers, and then after 3 convolutional layers, the same output result is obtained, which is the same as the sliding window result. However, using full convolutional layers makes the computation more efficient, as the former requires iterative computation while the latter can share resources.

![Framework of the weakly supervised recognition algorithm](image)

**Figure 1** Framework of the weakly supervised recognition algorithm

In building the model, the BN layer is added after the convolutional layer to improve the convergence speed and generalization ability, and the forward propagation formula of the BN layer is shown in equation (1).

$$
\begin{align*}
    u &= \frac{1}{m} \sum_{i=1}^{m} x_i^2 \\
    \delta^2 &= \frac{1}{m} \sum_{i=1}^{m} (x_i^2 - u)^2 \\
    x_i &= \frac{x_i - u}{\sqrt{\delta^2 + \epsilon^2}} \\
    y_i &= \gamma \times x_i^2 - \beta
\end{align*}
$$

In classification tasks, neural network-based algorithms generally add fully connected layers after convolutional layers to integrate the extracted features in vectorized form. Sometimes several fully connected layers need to be designed to improve the classification performance. However, this approach has significant drawbacks, on the one hand, fully connected layers multiply the network parameters and increase the training cost; on the other hand, fully connected layers can destroy the image spatial structure and are not conducive to the extraction of location information. In this paper, the fully connected layer is replaced by a convolutional layer to circumvent these drawbacks. Then the global average pooling layer is connected after the convolutional layer, which is reflected in the down sampling, retaining the significant object feature information while reducing the feature dimensionality. The purpose of the global average pooling layer is to output a separate image-level score for each object class and the size of the input image can be variable.

Weakly supervised learning is theoretically feasible because of the intrinsic correlation between whole-image classification and object localization/detection. For example, an image is labelled as "bird" because there is indeed a bird in this image. Although it is not certain where the bird is, the information derived from the area where the bird is located is better at identifying the category of the
whole image than the background.

### 3.2. Experimental design
Throughout the training process, the generator has to run on the region where only the attention mask is added, so that as the attention network is trained, the foreground part of the original image becomes clearer and the generator transforms only the region that needs to make changes between the source and target domains. However, the existing problem is that the generator only focuses on the generated part, while the discriminator focuses on the whole image, then it may have the consequence that the converted horse is a zebra, while the background part cannot correspond to it. In this paper, we do not use the generated image directly, but use the synthesized image as the input of the discriminator, which means that its distribution is neither the distribution of the source domain image nor the distribution of the target domain image. The generator will gradually draw the region of the background into the attention mask region as well, while the masked map will try to expand its attention region during the training process, with the consequence that all the attention mask maps converge to 1, as shown in Figure 2.

#### Data augmentation
Data augmentation is a common method for sample set expansion. Deep neural networks require a large amount of data to train the model to support many of the parameters required by the network, however, in practice we do not have as much data as we would like. Besides, if the target is to be applied to a variable environment, different positions, angles, and orientations of the objects will have an impact on the results, and it is often necessary to train the neural network by combining processed data to better handle the impact of the variable environment on the model. For example, if you want to train an image classifier about cats, if the prepared data samples are all from similar locations or the same angle, the classifier will be insensitive to cats from other angles or locations, i.e., it will not achieve better recognition results. A simpler and more efficient processing method is to flip and pan the images in the existing dataset, and then get the desired results when training the network with the expanded dataset [8]. Such a processing method has a major premise that the convolutional neural network is invariant, i.e., it remains invariant to size, translation, and rotation, and only extracts the essential shallow information in the image.

To address this problem, this paper attempts to minimize the difference between the relevant parts of the data generating distribution in the source and target domains. These maps are then applied back to the inputs of the generators to constrain them to the relevant image regions. No additional supervised training is required for the entire network model. It is shown through qualitative and quantitative analysis that explicitly incorporating this recognition mechanism into an image translation network can significantly improve the quality of translated images.
4. Result Analysis

In this paper, several sets of comparison experiments are conducted based on the supervised approach of image-level labelling, and the experiments are built on the VGGNet model. The model is first pre-trained on the ImageNet classification dataset so that the pre-trained model has the ability of feature extraction, and then the weights of the feature extraction convolutional layer are retained, and the model is trained and tested on the Pascal VOC dataset. In the first set of experiments, only the final fully-connected classification layer is changed to multi-label classification, and the other parts are not modified to train the fully-connected classification layer and complete the test, called VGGNet; in the second set of experiments, all the original fully-connected layers on the VGG model are replaced with convolutional layers, and then the global maximum pooling layer and classification layer are added, and the new convolutional adaptation layer, pooling layer, and classification layer are retrained based on retaining the weights of the original convolutional layer and testing the performance, called VGGNet. The third group of experiments replaces only the global maximum pooling layer with the global average pooling layer based on the second group and tests the performance, called VGG-average.

![Figure 3](image-url) Comparison algorithm of AP

Figure 3 shows the classification accuracy of these models on the VOC dataset. The fully connected layer destroys the localization ability of the network and affects the classification ability of the model. In terms of classification performance, the results of the second and third sets of experiments are not very different, but the third set of experiments is the best, with more outstanding performance in AP metric.

![Figure 4](image-url) Running time of label estimation for eight graph-based weakly supervised learning methods with different number of training samples
Figure 4 shows the computational complexity of different graph-based semi-supervised learning methods. These synthetic datasets consist of ten Gaussian clusters whose centres are sampled from the vertices of the Latin hypercube. The experiments are divided into two settings: a fixed number of feature dimensions of 10 for a different number of training samples, and a fixed number of training samples of 10,000 for a different number of feature dimensions.

Finally, the visualization method of the convolutional neural network is used to obtain the approximate location information of the objects by weighted summation of the convolutional feature maps. The model training and detection under the natural scene’s dataset are realized, and the weakly supervised object recognition task is completed.

5. Conclusion
Recognition tasks in computer vision, image classification, target object detection, and semantic segmentation of images, achieve promising results in supervised machine learning frameworks. However, for such machine learning systems to be widely used in real-world tasks with good performance and robustness, they need to rely on a large amount of accurately labelled training data. Obtaining large amounts of accurately labelled training data is costly in terms of time and labour. Exploring machine learning methods for the case of only a small number of annotations can help reduce the time and labour costs required to obtain accurately annotated samples. Therefore, in this paper, we focus on three major recognition tasks in the field of computer vision and investigate their corresponding finite-sample machine learning algorithms, i.e., non-fully supervised machine learning algorithms, and propose improvements and perform experimental validation. In this paper, we propose new and improved algorithms for the three major tasks in computer vision, image classification, target detection, and semantic segmentation of images, in terms of their corresponding semi-supervised learning classification algorithms, weakly supervised target localization algorithms, and few-sample semantic segmentation algorithms, respectively, around the topic of finite-sample supervised learning.

References
[1] Campanella G, Hanna MG, Geneslaw L, Miraflor A, Silva VW, Busam KJ, Brogi E, Reuter VE, Klimstra DS, Fuchs TJ. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature medicine, vol. 25, no. 8, pp. 1301-1309, 2019.
[2] Caicedo JC, Cooper S, Heigwer F, Warchal S, Qiu P, Molnar C, Vasilevich AS, Barry JD, Bansal HS, Kraus O, Wawer M. Data-analysis strategies for image-based cell profiling. Nature methods, vol. 14, no. 9, pp. 849-863, 2017.
[3] Smith K, Piccinini F, Balassa T, Koos K, Danka T, Azizpour H, Horvath P. Phenotypic image analysis software tools for exploring and understanding big image data from cell-based assays. Cell systems, vol. 6, no. 6, pp. 636-653, 2018.
[4] Wagholikar KB, Estiri H, Murphy M, Murphy SN. Polar labeling: silver standard algorithm for training disease classifiers. Bioinformatics, vol. 36, no. 10, pp. 3200-3206, 2020.
[5] Yang X, Wang Y, Byrne R, Schneider G, Yang S. Concepts of artificial intelligence for computer-assisted drug discovery. Chemical reviews, vol. 119, no. 18, pp. 10520-10594, 2019.
[6] Rundo L, Militello C, Vitabile S, Russo G, Sala E, Gilardi MC. A survey on nature-inspired medical image analysis: a step further in biomedical data integration. Fundamenta Informaticae, vol. 171, no. 1-4, pp. 345-365, 2020.
[7] Yan J, Wang C, Cheng W, Gao M, Zhou A. A retrospective of knowledge graphs. Frontiers of Computer Science, vol. 12, no. 1, pp. 55-74, 2018.
[8] Patrascu MB, Pottel J, Pinus S, Bezanson M, Norrby PO, Moitessier N. From desktop to benchtop with automated computational workflows for computer-aided design in asymmetric catalysis. Nature Catalysis, vol. 3, no. 7, pp. 574-784, 2020.