Hierarchical Feature Mining for Image Classification and Segmentation

Hayder Yousif[1] and Zahraa Al-Milaji[2]
{hayder.yaqoob@stu.edu.iq[1], zahraa.a.nejim@stu.edu.iq[2]}

Basra Engineering Technical College, Southern Technical University, Iraq[1,2]

Abstract. To identify and segment two different types of images, supervised and unsupervised learning are utilised. First, we design a fast and accurate feature extraction strategy using our deep learning classification model. We create a DCNN model that is 18 times faster than the current state-of-the-art AlexNet model while sacrificing only a little amount of accuracy. Second, several classifiers have been trained on the characteristics retrieved from the deep convolutional neural networks' fully connected layer (DCNN). Finally, we perform K-means graph-cut segmentation and compare the results to those of supervised segmentation. Natural scenes and biomedical image processing were used to investigate image-level and region-level classification. On a biomedical picture dataset, our experimental results show that the suggested strategy outperforms existing methods..

Keywords: DCNN, supervised learning, unsupervised learning, K-means, graph-cut

1 Introduction

In intelligent image analysis, classifying and segmenting items from an image is a crucial and enabling step. DCNN [1, 2] has recently made significant contributions to a variety of computer vision challenges, including object detection and semantic segmentation. By creating a graph and computing the smallest cut, graph cut methods tackle energy reduction challenges. The minimal cut challenge aims to locate a cut with a capacity equal to the graph's fewest overall cuts [3]. The minimum is found by calculating the sum of all cuts that divide the graph into two nonempty pieces [4]. Subspace segmentation can be done as a data grouping issue by first learning an affinity matrix from the input data, and then using spectral clustering methods like Normalized Cuts (NCut) to get the final segmentation results [5, 6]. The vast number of social networks dedicated to the sharing of photographs of cats and dogs [7] suggests that people care deeply about their domestic animals. Object category recognition and segmentation have been integrated in various ways in the past. However, segmentation of the entire image has been a common goal of various approaches [8].

Recent studies have worked on epithelium and stroma on the TMAs samples. Epi-stroma dataset [9] contains two annotated image classes in which each the image corresponds to the presence of either epithelium or stroma collected from

IMDC-IST 2021, September 07-09, Sakarya, Turkey

DOI 10.4108/eai.7-9-2021.2315306

643 patients tissue samples with colorectal cancer. [10] extracted morphological features from super pixels of breast cancer tissue samples. Bianconi et al [11] used perceptual feature space to classify image blocks into tumour epithelium and stroma classes. Linder et al. [9] have performed texture analysis for images divided into rectangular regular blocks in digitized tumour tissue microarrays (TMAs). AiPing et al [12] have extracted pixel-level colour and texture features to identify epithelial carcinoma tissue from stromal tissue.
In this paper, two image analysis tasks have been studied. In the first task, DCNN model which is an application of deep learning was used for automated feature extraction. Then, different classifiers are evaluated using different metrics. In the second task, graph-cut is used to separate the image regions based on their spatial connectivity.

2 Datasets

We employed two datasets in this study, each with its own set of image gathering settings. The Oxford-IIIT Pet dataset, for example, contains 7,349 photos of cats and dogs of 37 different breeds, 25 of which are dogs and 12 of which are cats. This dataset is separated into two parts: a training set with 3680 photos, and a testing set with 3669 images. Figure 1 depicts samples from this collection.

Fig. 1. Oxford-IIIT Pet dataset samples.

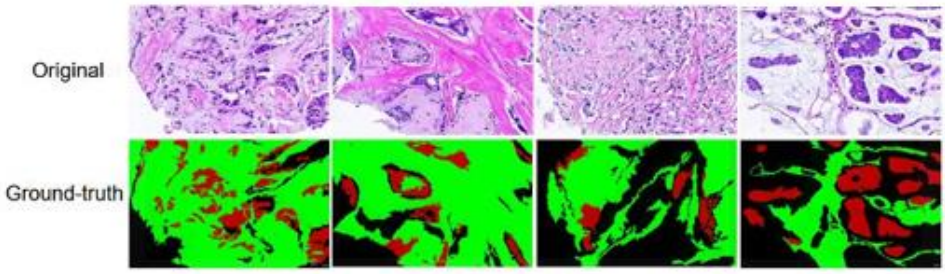

Fig. 2. Samples from the VGH dataset.

The epi-stroma dataset from the Vancouver General Hospital (VGH) is the second dataset. Pathologists manually annotated epithelial and stromal areas in this dataset. Only 36 images are used for training and 15 images are used for testing in this example. Figure 2 shows samples from the VGH dataset. Green represents the stromal regions, whereas red represents the epithelial regions. The pathologists have not classified the dark areas.

3 Region Extraction and Pre-processing for the VGH Images

For the VGH dataset, we divide the image into 96 × 96 overlapped regions in both training and testing steps. At this point, we need to extract a pure single class region for more accurate training without overfitting. We propose to purify the mixed class regions and replace the non-classified by a pathologist (black ground truth regions) by white pixel. The white pixels region has no texture information that may cause information corruption. Fig. 3 shows some refined regions.

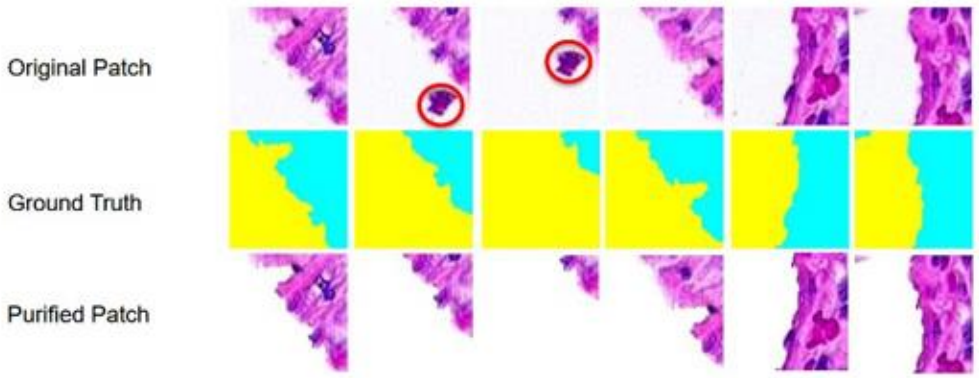

Fig. 3. Mixed class region purifying in the VGH dataset.

4 DCNN for Feature Extraction

In the DCNN, we investigate the complexity-accuracy trade-off. The DCNN, as we all know, is a computationally expensive algorithm. As a result, there is a pressing need to accelerate the DCNN while retaining its classification accuracy. The input size, number of layers, and number of filters are three important parameters that can be used to efficiently regulate the computational complexity of DCNN. The 14 levels of our developed DCNN architecture are shown in Table 1. Rectified Linear Unit ReLU layer follows each convolutional layer, while normalisation and pooling layers follow the ReLU layer.

Table 1. Our CNN networks architecture.

Layer	Type	K size	Filter #	Stride	Pad
1	Conv1	6	128	2	0
4	Pool1	3	2	0	
5	Conv2	5	128	1	2
8	Pool2	5	2	0	
9	Conv3	5	128	1	1
11	FC4	7	2048	1	0

Using FC5, we create a 2048-dimensional feature vector from each region proposal. We use the linear SVM, SoftMax, KNN, and Random-Forest trained for that class to score the extracted feature vector for each class. Table 2 shows the resource needs and testing speed related with various DCNN systems. We can see that constructing smaller input images (9696 pixels) with fewer layers reduces complexity by 18 times while sacrificing classification accuracy.

Table 2. DCNN Training Memory allocation and patch testing time comparison.

DCNN	Memory(MB)	(patch/ms)
AlexNet [1]	217	336.1
VGG-F [13]	316	395.6
VGG-S [13]	393	304.7
ours	50	17.9

5 Classifiers

Here, we briefly describe the functionality of the used classifiers.

5.1 SoftMax

SoftMax regression (also known as multinomial logistic regression) is a multiclass expansion of logistic regression. The labels in logistic regression are considered to be binary: $y^{(i)} \in \{0, 1\}$. Remember that in logistic regression, we had a training set of m labelled samples (x(1), y(1)),..., (x(m), y(m)), where the input features were $x^{(i)} \in \Re^n$. In the binary classification context of logistic regression, the labels were y(i) 0 and 1. The hypothesis was as follows:

$$h_\theta(x) = \frac{1}{1 + \exp(-\theta^\top x)}, \tag{1}$$

$$J(\theta) = -\left[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))\right] \tag{2}$$

6 SVM

SVM has grown in popularity as a method for handling classification, regression, and novelty detection problems. The determination of the model parameters corresponds to a convex optimization problem, therefore any local solution is also a global optimum, which is an important property of support vector machines.

The decision boundary in support vector machines is chosen to be the one for which the margin is maximized. SVM is a kernel-based algorithm that has sparse solutions so that predictions for new inputs depend only on the kernel function evaluated at a subset of the training data points. Our goal is now to maximize the margin between classes while softly penalizing points that lie on the incorrect side of the margin boundary.

6.1 Random-Forest

Random forest picks a subset of features at random to build a fully developed decision tree on the n data points, using the sub-feature set as a replacement sample. Then repeat these two processes to make a random forest with many trees. Apply the decision to test data for each tree in the forest, then use the majority vote of all trees to make the final prediction.

6.2 K-Nearest Neighborhood

KNN is one of the nonparametric approaches that place very mild assumptions on the data distribution and provides good models for complex data [13]. For example, inserting a cell about x and letting it grow until it captures kn samples, where kn is some predetermined function of n, can be used to estimate p(x) given n training samples. These are x [14's kn nearest-neighbors samples. The distribution of all its class data points determines the distance measurement type and number of neighbours for a given data point. We used cosine distance for both datasets. SNPHEp-2 and Epistomes were chosen by a majority vote among 7 and 3 local points, respectively.

7 Graph cut for Unsupervised Image Segmentation

G = V, E > is a directed weighted graph with a set of nodes V and directed edges E connecting them.

$$D_{xyk} = \| I_{xyk} - C_{xyk} \|^2 \quad (3)$$

Pixels, voxels, and other features are usually represented by nodes. Classes are additional special nodes seen in some graphs [15]. The edges of all graphs are given a weight or cost. N-links and T-links are the two types of graph edges. N-links join adjacent pixels or voxels together. As a result, in the image, they represent a neighbourhood system.

The cost of N-links is a penalty for pixels that are not connected. The pixel interaction term Vx,y in energy is used to calculate these expenses. Pixels with class are connected by T-links. A penalty for assigning the corresponding class to a pixel relates to the cost of a T-link connecting a pixel and a class. The connectedness of each pixel value to the class prototype is measured using Euclidean distance.:

We used a square matrix M with K rows and columns to represent the cost. When the matrix size is increased, the neighbourhood constraints are strengthened even more and the segments become larger. To produce the final segmentation result, graph cut is done to the distance D with the cost M.

8 Experimental Results

8.1 Quantitative Results

To compute the classification outputs, we use the following performance metrics:

$$Recall = \frac{TP}{TP + FN}, \tag{4}$$

$$Precision = \frac{TP}{TP + FP}, \tag{5}$$

$$F - measure = \frac{2 \times Recall \times Precision}{Recall + Precision}. \tag{6}$$

The total number of true positives, false positives, and false negatives are represented by the letters TP, FP, and FN, respectively. Note that the F-measure needs to be as high as possible, to minimize the classification errors. In Table 3, we evaluate the classification accurateness of various classifiers on Oxford-III pet dataset. Because the required square input image needs to be fed to the DCNN, we rescale the different aspect ratio input images to fit with the required size. This corrupts the spatial information of the cats and dog effect the discrimination decision. The division of the VGH dataset into overlapping blocks cannot generate perfect epithelial-stromal regions but it is still a reliable solution even with non-smooth blocky output. Our pre-processing of the training patches (regular blocks extracted from an image) helps to improve the classification accuracy to outperform the most recent work [16] on the VGH images.

Table 3. Performance comparison on Oxford-III Pet dataset using different classifiers.

Classifier	Accuracy %	Precision %	Recall %	F-measure %
SoftMax	81.11	64.67	73.56	68.83
SVM	81.08	62.89	74.47	68.19
Random-Forest	80.81	64.16	73.05	68.32
KNN	78.36	39.31	86.95	53.94

Table 4. Performance comparison on Oxford-III Pet dataset using different classifiers.

Classifier	Accuracy %	Precision %	Recall %	F-measure %
SoftMax	88.99	88.89	97.48	92.98
SVM	88.92	88.81	97.43	92.92
Random-Forest	88.11	87.6	97.88	92.46
KNN	84.96	83.77	98.97	90.74
Xu et. al [16]	88.34			

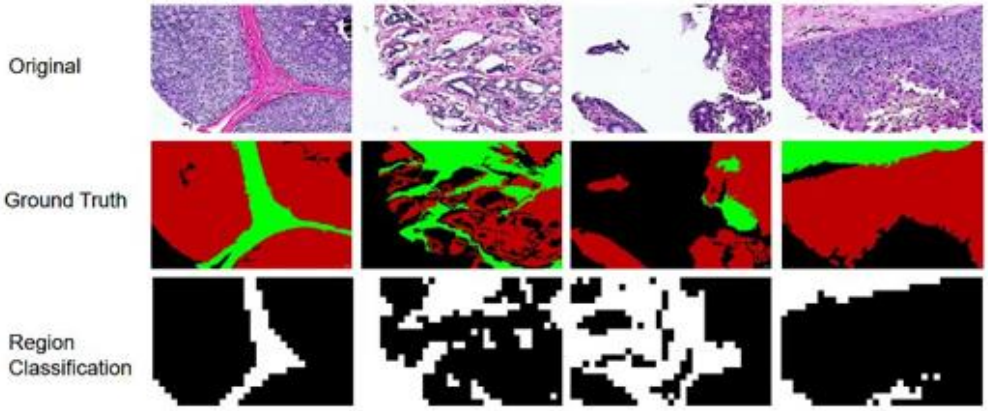

Fig. 4. VGH dataset classification results.

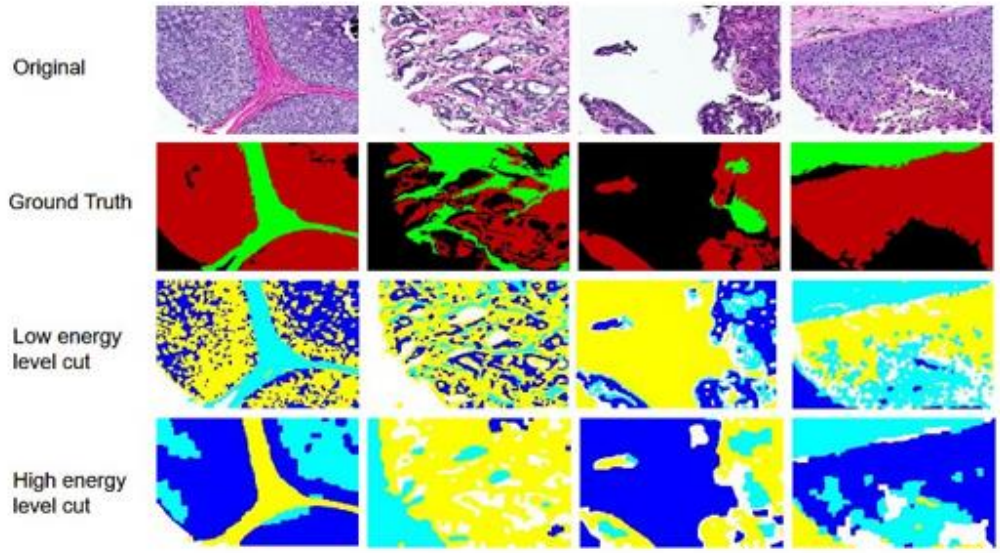

Fig. 5. VGH dataset graph-cut segmentation results.

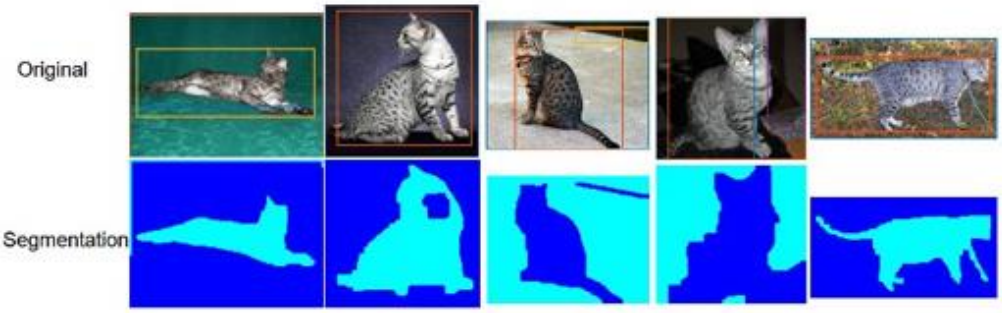

Fig. 6. Oxford-III pet dataset graph-cut segmentation results.

8.2 Qualitative Results

Fig. 4 shows the qualitative results after concatenating the classification labels of the SoftMax to fit with their locations in the original image. Stromal regions are in white colour on the classification output are corresponding to green regions in the ground truth image. Epithelial regions are in black on the classification output is corresponding to red regions in the ground truth image.

We also tried to segment the VGH images using an unsupervised method called graph cut. From Fig. 5, we can see the smooth regions produced by this method as compared with the supervised method used in Fig. 4. However, in the un-supervised learning the class label is not known for the regions. It has been tried two levels of energy minimization to illustrate its effect on the region size and smoothness. We used the graph cut to separate a pet from the background to avoid training the classifier on noisy features. Fig. 6 shows some samples of the segmentation output. Again, we can see only two regions in the output but no label for that region to tell the difference between foreground and backdrop regions.

9 Conclusion

We used the DCNN as an automated method for feature extraction instead of using handcrafted features like HOG and LBP. Our experiments show that SoftMax outperforms the other classifiers. Mixed class purifying increases the classification accuracy to outperform VGH data. For pet data, rescaling different aspect ratio images to fit with square input image for DCNN decreases the classification accuracy. We also tried to segment the image by applying constraints to the K-means using graph-cut. Changing the level of energy controls the shape and the size of the segmentation regions. Fusion of the supervised and unsupervised method may increase the detection/segmentation performance.

References

[1] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems; 2012. p. 1097-105.

[2] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998;86(11):2278-324.

[3] Tao W, Tai XC. Multiple piecewise constant with geodesic active contours (MPC-GAC) framework for interactive image segmentation using graph cut optimization. Image and Vision Computing. 2011;29(8):499-508.

[4] Wu Z, Leahy R. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. IEEE transactions on pattern analysis and machine intelligence. 1993;15(11):1101-13.

[5] Shi J, Malik J. Normalized cuts and image segmentation. IEEE Transactions on pattern analysis and machine intelligence. 2000;22(8):888-905.

[6] Liu G, Lin Z, Yan S, Sun J, Yu Y, Ma Y. Robust recovery of subspace structures by low-rank representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013;35(1):171-84.

[7] Parkhi OM, Vedaldi A, Zisserman A, Jawahar C. Cats and dogs. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE; 2012. p. 3498-505.

[8] Parkhi OM, Vedaldi A, Jawahar C, Zisserman A. The truth about cats and dogs. In: Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE; 2011. p. 1427-34.

[9] Linder N, Konsti J, Turkki R, Rahtu E, Lundin M, Nordling S, et al. Identification of tumor epithelium and stroma in tissue microarrays using texture analysis. Diagn Pathol. 012;7(22):1596-7.

[10] Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, van de Vijver MJ, et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Science translational medicine. 2011;3(108):108ra113-3.

[11] Bianconi F, Álvarez-Larrán A, Fernández A. Discrimination between tumour epithelium and stroma via perception-based features. Neurocomputing. 2015;154:119-26.

[12] Qu A, Chen J, Wang L, Yuan J, Yang F, Xiang Q, et al. Two-step segmentation of Hematoxylin-Eosin stained histopathological images for prognosis of breast cancer. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2014. p. 218-23.

[13] Vedaldi A, Lenc K. Matconvnet: Convolutional neural networks for matlab. In: Proceedings of the 23rd ACM international conference on Multimedia. ACM; 2015. p. 689-92.

[14] Duda H, Hart P, et al.. Stork, Pattern Classification. John Wiley & Sons; 2001.

[15] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. IEEE Transactions on pattern analysis and machine intelligence. 2001;23(11):1222-39.

[16] Xu J, Luo X, Wang G, Gilmore H, Madabhushi A. A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. Neurocomputing. 2016;191:214-23.