Active Learning for Patch-Based Digital Pathology using Convolutional Neural Networks to Reduce Annotation Costs

Carse, Jacob; McKenna, Stephen

Published in:
Digital Pathology

DOI:
10.1007/978-3-030-23937-4_3

Publication date:
2019

Document Version
Peer reviewed version

Active Learning for Patch-Based Digital Pathology using Convolutional Neural Networks to Reduce Annotation Costs

Carse, Jacob; McKenna, Stephen

Published in:
Digital Pathology

DOI:
10.1007/978-3-030-23937-4_3

Publication date:
2019

Document Version
Peer reviewed version

Active Learning for Patch-Based Digital Pathology using Convolutional Neural Networks to Reduce Annotation Costs

Jacob Carse and Stephen McKenna

CVIP, School of Science and Engineering, University of Dundee, Dundee DD1 4HN, Scotland, UK
{j.carse, s.j.z.mckenna}@dundee.ac.uk

Abstract. Methods to reduce the need for costly data annotations become increasingly important as deep learning gains popularity in medical image analysis and digital pathology. Active learning is an appealing approach that can reduce the amount of annotated data needed to train machine learning models but traditional active learning strategies do not always work well with deep learning. In patch-based machine learning systems, active learning methods typically request annotations for small individual patches which can be tedious and costly for the annotator who needs to rely on visual context for the patches. We propose an active learning framework that selects regions for annotation that are built up of several patches, which should increase annotation throughput. The framework was evaluated with several query strategies on the task of nuclei classification. Convolutional neural networks were trained on small patches, each containing a single nucleus. Traditional query strategies performed worse than random sampling. A K-centre sampling strategy showed a modest gain. Further investigation is needed in order to achieve significant performance gains using deep active learning for this task.

Keywords: Active learning · Image annotation · Deep learning · Nuclei Classification

1 Introduction

Modern deep learning algorithms have been shown to improve performance for tasks such as classification, segmentation and detection in digital pathology. However, deep learning algorithms require large annotated datasets from which to build high performing models. This requirement for data has been identified as a key challenge for using deep learning algorithms for digital pathology [12] and medical image analysis [5]. There are several approaches to tackling this problem which include semi-supervised learning, weakly supervised learning, active learning, and their combinations. This paper focuses on the use of active learning to aid in annotation collection for patch-based digital pathology image analysis.

Active learning is a type of machine learning which hypothesises that having a learning algorithm select the data it uses to train itself can reduce the amount of data needed for training. Active learning is used within modern applications to reduce the quantity of annotations needed. Annotating only the data selected by the learning algorithm reduces the overall cost of building an effective model. In a pool-based scenario, the learning algorithm has access to a large pool of unannotated data. Over multiple iterations, the learning algorithm selects data to be annotated and added to the training data [9].

Active learning algorithms use query strategies to select the data to be annotated. There are numerous query strategies available with the most popular methods being based on uncertainty. Uncertainty sampling is a simple query strategy that samples data based on a models predictions for the unannotated data.

While these methods have been shown to work well with many traditional learning algorithms, this is not the case when working with deep learning algorithms. There are several reasons for this. Firstly, deep learning algorithms jointly learn feature representations and classifiers/regressors. Selecting only difficult examples to train the model leads to learnt features that are not representative, decreasing the quality of the model [13]. Secondly, traditional query strategies are used to select a single data point. It has been shown that deep learning algorithms work better with batch updates and so require a query strategy to select an optimal batch and not just the top ranked points [8]. Thirdly, the softmax output from a classifier trained using deep learning algorithms does not represent the models uncertainty well, which is commonly used to sample unannotated data [4].

A standard approach for using learning algorithms for digital pathology is to use patches from larger images. This allows the images to be input to learning algorithms such as convolutional neural networks (CNNs) more efficiently and removes the necessity of annotating very large images. When using patch-based methods that use small patches, for tasks such as nuclei detection and classification, using active learning to select patches for annotation can be detrimental for annotation collection. This is due to small patches being time consuming and tedious to annotate. Small patches may also not include enough context for accurate annotation.

To address these issues, this paper modifies query strategies so that tasks which rely on small patches can efficiently use active learning and ease the effort needed from expert annotators. Methods were tested using the CRCHistoPhenotypes dataset for nuclei detection and classification [10].

2 Related Work

Since the rise in popularity of deep learning, numerous active learning query strategies have been proposed. Some are simple alterations to existing active learning query methods, such as Cost-Effective Active Learning (CEAL) [13]. In this algorithm, predictions are made for all the unannotated data and a batch

is then selected from the data with the highest uncertainty. In addition to this, the most confident predictions are also added to the training data with their predicted label. This increases the overall size of the training dataset without any extra annotation cost, adding data which is easier to classify so that the model can learn more representative features during training. Other methods seek to alter the deep learning algorithms so that traditional active learning query strategies can be used. Using Bayesian deep learning algorithms to produce more accurate uncertainty metrics is an example of this [4].

The Core-Set query strategy [8] focuses on selecting a batch of unannotated data that can be used for both learning representative features and optimising the classifier. This is done by treating the problem as a cover problem and using a mixed integer programming heuristic to minimise the covering radius of the data. Another query strategy for deep learning uses adversarial attacks to estimate the decision boundaries of the model and selects the data closest to the decision margins [3]. These query strategies achieve state of the art results when working with CNNs, demonstrating how active learning has potential for working with deep learning algorithms.

The popularity of deep learning inspired multiple applications in digital pathology [5]. These rely on availability of large annotated datasets such as the CRCHistoPhenotypes dataset [10]). Application of deep learning algorithms is limited by dataset availability. Despite this, numerous advances in digital pathology have been made using deep learning, for example in nuclei detection and classification [10], organ segmentation [1] and classification of diseases [7].

The expense of annotating large quantities of data has led researchers to investigate how active learning might be applied to digital pathology problems. Cosatto *et al* [2] used active learning to collect 10,000 nuclei annotations which were then used to train a machine learning model for nuclear grading. Yang *et al* [14] developed an active learning framework for digital pathology segmentation, specialised for that task.

3 Proposed Methods

Patch-based methods are common within digital pathology and medical image analysis more generally. However, applying active learning to these methods can be tedious, especially in systems that use small patches. Small patches can be difficult to annotate in isolation. Even if their spatial visual context in provided to the annotator, continually having to reassess context for each annotation can be inefficient and frustrating. We propose a region-based alternative that requests annotations over regions containing multiple small patches. Working with larger regions eases the effort needed from the annotator and can lead to an improved annotation collection throughput. This alteration allows for a learning algorithm to be trained with the small patches and only treats the data as regions when querying the unannotated data.

The proposed query strategy makes a simple modification to how an existing query strategy works. An overview of this can be seen in Algorithm 1 where S

is an existing query strategy. This algorithm is called at the end of each active iteration, once a model has been trained on the currently available annotations. It extracts all the patches from each unannotated region and make predictions on each patch. These predictions are then averaged to create a prediction for the overall region. Once all the regions have predictions, these predictions can be used within an active learning query strategy. An example of this would be using entropy uncertainty sampling where an uncertainty value for each region would be calculated and sampled. However, this approach can also be applied to more complex query strategies such as core-set sampling, by solving the K-centre problem for the region predictions rather than feature representations for individual data points.

4 Experiments

A nuclei classification task was chosen to investigate the effectiveness of using region-based active learning within digital pathology. This task used the CRCHistoPhenotypes dataset which consists of 22,444 annotated nuclei from 100 H&E stained histology images [10]. Coordinates for each nucleus along with their corresponding classifications have been annotated in this dataset. Each cropped histology image was split into 2,500 100x100 pixel regions from which 30x30 pixel patches were extracted for each nucleus. Augmentation was used during training, each patch being augmented by having a Gaussian blurring filter applied, and by horizontal and vertical flipping.

This experiment used a simple CNN inspired by the architecture used in the nuclei classification benchmark for the CRCHistoPhenotypes dataset [10]. It consisted of two convolutional layers, one with 36 filters with a size of 4x4 and the other with 48 filters with a size of 3x3, both of which were followed

Algorithm 1: Alteration to query strategy for region-based active learning

Input : θ are the trained weights for the learning algorithm,
δ is the learning algorithm,
U is the set of unannotated data,
n is the batch to be selected,
S is the query strategy that will be used.

Output: U' which is a sampled set from U

RegionQueryStrategy θ, δ, U, n, S
  foreach *region* r *in* U **do**
    $P \leftarrow \text{ExtractPatches}(r)$ — extract patches from region
    $O \leftarrow \delta(\theta, P)$ — makes predictions on extracted patches
    $O' \leftarrow \text{Average}(O)$ — average predictions
    $Y := Y + O'$ — append region average to array of averages
  end
  $U' \leftarrow S(Y, n)$ — select regions to query using the query strategy
  return U'

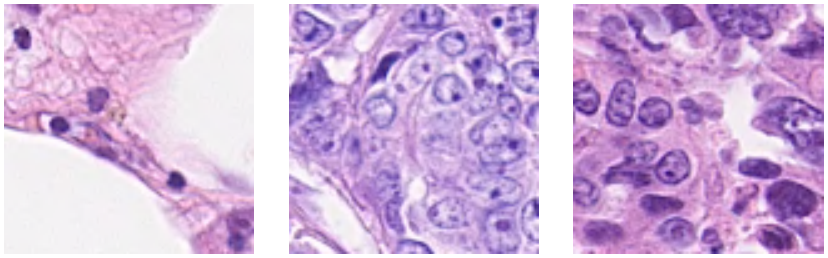

Fig. 2: Three example regions from the CRCHistoPhenotypes dataset [10] with multiple nuclei that will be extracted into patches and augmented.

by max pooling layers with a filter size of 2x2. These layers were followed by two fully connected layers with 1200 neurons and 512 neurons respectively. This architecture is summarised in Table 1. Each hidden layer used ReLU activation functions and the two fully connected layers used dropout for regularisation [11]. Dropout was also used to adapt the CNN into a Bayesian CNN.

The training environment is constantly changing between iterations as the dataset expands. The Adadelta training algorithm for gradient decent was chosen as the CNNs training optimiser [15]. Adadelta requires no manual tuning of learning rate as it adapts based on the training gradients, making it ideal for active learning tasks. To ensure that the model has been trained after each active iteration and that overfitting have been avoided, an early stopping method was used. The early stopping method chosen compares the generalisation loss (Equation 1) and training progression (Equation 2) and will stop training before overfitting [6]. Generalisation loss is calculated by comparing the validation loss for each epoch $L_{val}(t)$ against the minimum validation loss across all epochs. The training progression value is calculated by analysing the training losses $L_{tr}(t)$ over a batch of recent epochs of size k.

Convolutional Neural Network Architecture for Nuclei Classification		
Type	Filter Dimensions	Input/Output Dimensions
I		30 x 30 x 3
C	4 x 4 x 1 x 36	26 x 26 x 36
M	2 x 2	12 x 12 x 36
C	3 x 3 x 36 x 48	10 x 10 x 48
M	2 x 2	5 x 5 x 48
F	5 x 5 x 48 x 1200	1 x 1200
F	1 x 1 x 512 x 512	1 x 512
F	1 x 1 x 512 x 4	1 x 4

Table 1: The Convolutional Neural Network architecture for nuclei classification used in the region-based active learning experiments.

$$GL(t) = 100 \cdot \left(\frac{L_{va}(t)}{\underset{t' \leq t}{min} L_{va}(t')} - 1 \right) \tag{1}$$

$$P_k(t) = 1000 \cdot \left(\frac{\sum_{t'=t-k+1}^{t} L_{tr}(t')}{k \cdot min_{t'=t-k+1}^{t} L_{tr}(t')} - 1 \right) \tag{2}$$

Experiments tested the region-based modification combined with a range of query strategies. These query strategies included several more basic methods which will be used specifically to act as baselines for the other query strategies, built specifically for deep learning algorithms. These basic query strategies are random, least confident uncertainty, margin uncertainty and entropy uncertainty sampling. The other query strategies tried were K-Centre sampling (solved using a greedy approximation), Core-Set sampling [8] and Bayesian active learning by disagreement (BALD) sampling using Bayesian neural networks [4].

In each experiment, all available data were initially treated as unannotated; two randomly selected regions were then used to form the initial annotated training set. After each active iteration, two regions were selected from the unannotated regions to be added to the training set. This was continued for 50 iterations meaning that 102 regions out of 2,500 formed the final training set in each experiment. Each experimental setting was run five times with different random seeds (different weight initialisation and initial annotated patches).

5 Results

Table 2 gives the test accuracy and loss (averaged over the five runs) after 50 iterations for each of the query strategies. These results were obtained on a single, unchanging test set. Notably, only K-Centre sampling achieved a higher average accuracy than a random sampling strategy. Core-set sampling accuracy was very similar to that of random sampling. The other query strategies were all worse than simply adopting random sampling. Figs. 3 and 4 show the test accuracy and loss for each strategy after each iteration.

For comparison, a fully supervised CNN trained on a much larger training set of 2,500 annotated regions achieved an accuracy of 68.53% and a loss of 1.111. Training using the K-Centre query strategy achieved an accuracy of 61.41% and a loss of 1.137 using 4% of the annotations.

Query Strategy	Random	Least Confident	Margin	Entropy	K-Centre	Core-Set	BALD
Accuracy	58.25%	48.92%	45.84%	32.37%	61.41%	57.33%	48.23%
Loss	1.154	1.243	1.268	1.39	1.123	1.157	1.247

Table 2: The accuracy and loss for each model trained with different query strategies over 50 iterations resulting in a total of 102 annotated regions.

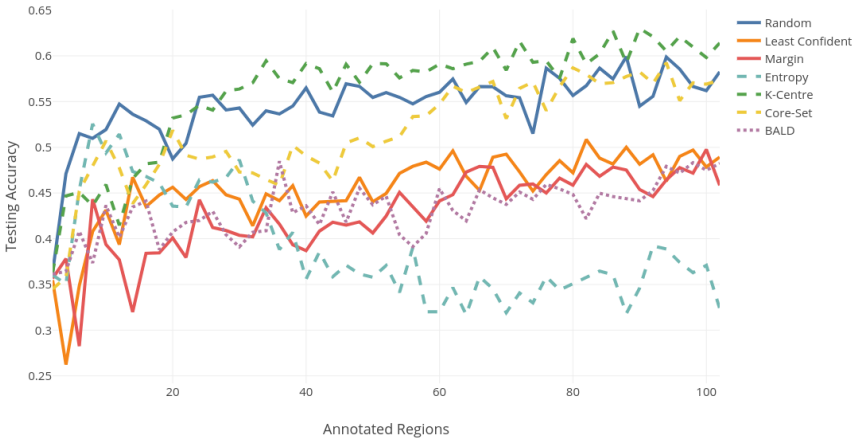

Fig. 3: Test accuracy across active iterations.

Fig. 4: Test loss across active iterations.

6 Conclusion

This paper proposed a mechanism for applying deep active learning to patch-based systems with specific focus on its application to nuclei classification. The results clearly showed that the traditional active learning query strategies performed poorly. Active learning methods tailored to deep CNNs are needed. Reducing annotation overheads and thus the cost of developing deep learning systems for digital pathology and medical image analysis can allow those with less

access to resources to work on a range of problems. Methods such as active learning have great potential but further work is needed in order to achieve significant gains on tasks such as that investigated here.

References

1. BenTaieb, A. and Hamarneh, G.: Topology Aware Fully Convolutional Networks for Histology Gland Segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 460-468 (2016)
2. Cosatto, E., Miller, M., Graf, H.P., Meyer, J.S.: Grading Nuclear Pleomorphism on Histological Micrographs. In: 19th International Conference on Pattern Recognition. pp. 1-4 (2008)
3. Ducoffe, M., Precioso, F.: Adversarial Active Learning for Deep Networks: a Margin Based Approach. In: Proceeding of Machine Learning Research **80** (2018)
4. Gal, Y., Islam, R., Ghahramani, Z.: Deep Bayesian Active Learning with Image Data. In: Proceedings of Machine Learning Research **70** 1183-1192 (2017)
5. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van DerLaak, J.W.M., Van Ginneken, B., Snchez, C.I.: A Survey on Deep Learning in Medical Image Analysis. In: Medical Image Analysis **42** 60-88 (2017)
6. Prechelt, L.: Early Stopping - But When?. In: Neural Networks: Tricks of the Trade 55-69 (1998)
7. Schaumberg, A.J., Rubin, M.A. and Fuchs, T.J.: H&E-Stained Whole Slide Image Deep Learning Predicts SPOP Mutation State in Prostate Cancer. In: BioRxiv BioRxiv:064279 (2018)
8. Sener, O., Savarese, S.: Active Learning for Convolutional Neural Networks: A Core-Set Approach. In: International Conference on Learning Representations (2018)
9. Settles, B.: Active Learning, In: Synthesis Lectures on Artificial Intelligence and Machine Learning **6**(1), 1–114 (2012)
10. Sirinukunwattana, K., Raza, S.E.A, Tsang, Y.W., Snead, D.R.J., Cree, I.A., Rajpoot, N.M.: Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images. In: IEEE Transactions on Medical Imaging **35**(5), 1196-1206 (2016)
11. Srivastava, N., Hinton, G., Krizevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A Simple Way to Precent Neural Networks from Overfitting. In: The Journal of Machine Learning Research, **15**(1) 1929-1958 (2014)
12. Tizhoosh, H.R., Pantanowitz, L.: Artificial intelligence and digital pathology: Challenges and opportunities. In: Journal of Pathology Informatics **9** (2018)
13. Wang, K., Zhang, D., Li, Y., Zhang, R., Lin, L.: Cost-Effective Active Learning for Deep Image Classification. In: IEEE Transactions on Circuits and Systems for Video Technology **27**(12), 2591-2600 (2017)
14. Yang, L., Zhang, Y., Chen, J., Zhang, S. and Chen, D.Z.: Suggestive Annotation: A Deep Active Learning Framework for Biomedical Image Segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 399-407 (2017)
15. Zeiler, M.: ADADELTA: An Adaptive Learning Rate Method. In:arXiv preprint arXiv:1212.5701 (2012).