Classifiying Transmission Electron Microscopy Virus Textures

Alan Do-Omri 260532985
Kian Kenyon-Dean 260564475
Genevieve Fried 260564432

December 9, 2015

Abstract

Determining the identity of a virus in a blood sample is a critical problem for today's medical community. Incorrectly classifying a virus can have grave implications, not only for the patient in question but also for those who've come into contact with him or her. A virus, even after microscopy techniques like the transmission electron microscopy method have been applied, can still be difficult to discern by the human eye. Unfortunately, this is not a problem that can afford a margin of error. Fortunately, machine learning methods present exceedingly high accuracy rates for image classification tasks and don't require expert knowledge to do so. Here we present a successful application of convolutional neural networks to the problem of virus classification. Using the Virus Texture Dataset from Uppsala University, we classify viruses samples into one of 15 categories and compare its performance to previous work done on the same task.

1. Introduction

2. Dataset

3. Methods

Convolutional Neural Networks (CNN) are biologically inspired variants of Multi-Layered Perceptrons that can encode non-linear features with atleast one hidden layer, and are specifically architected for image classification. Image processing with a Neural Network is computationally infeasible given how quickly parameters scale. CNN's on the other hand constrain their architecture to the nature of images. A neuron in the nth layer is not fully connected to all neurons before it, rather,

They work as follows.

They're constrained to work with images, which specifically oriented towards image processi Image processing with a fully connected Neural Network

Unlike a fully connectd Neural Network, CNN's a Neural Network, image classification is infeasibly computationally expensive, and CNN's CNN's are made up of several layers: input, convolutional, pooling and fully connected.

4. Related Work

In the work of Kylberg et al. [4], texture analysis is performed on a dataset composed of 22 different virus samples and the resulting feature vector is fed into a Random Forest classifier. The authors first compare the performance of different texture analysers, such as Local Binary Patterns, Radial Density Profile and their respective variants before proceeding to detail their results. We will briefly explain their feature extractors and results here.

4.1. Local Binary Profile (LBP)

Local Binary Profile works such that, given an image, for each pixel p_i in it we sample n equally-spaced points on the circle of radius r with center p_i – an example of sampling is shown in figure 1 – and con-

struct a vector $v(p_i)$ such that its ith entry is a 1 if the ith sampled pixel has a value bigger than p_i and 0 otherwise.

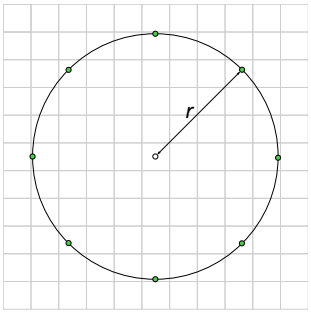


Figure 1: Example of LBP sampling: The green points are the neighbouring sample points at distance r from the central white point. In this case, we are sampling with $n = 8$ neighbour points.

A sequence of 0s and 1s are constructed $v(p_i)$ to form a binary number v_{p_i}, hence the name. Once we have the v_{p_i} for all pixels, we construct a histogram counting the number of appearances of each value v_{p_i}. The histogram forms the feature vector associated with the given image. Kylberg et al. [4] denote this feature extraction method by $\text{LBP}_{n,r}$, where n is the number of sampled points and r is the radius, as described above. The resulting histogram is represented in a vector of counts with 2^n elements.

Rotational Invariant LBP In order to reduce the size of the feature vector, Kylberg et al. [4] mention a modification of LBP in the following sense: instead of creating v_{p_i} by interpreting the vector $v(p_i)$ and using v_{p_i} as is, rotate the number v_{p_i} bitwise until you get the smallest possible number. For example, the number 110 becomes 011. They name this technique the rotational invariant and denote it with $\text{LBP}^{\text{ri}}_{n,r}$.

Uniform LBP To further reduce the size of the histogram, Kylberg et. al also restrict the values of v_{p_i} to only numbers that have 2 or less transitions from 0 to 1 or from 1 to 0, and they call this variant *uniform binary patterns with at most 2 spatial transitions*, denoted $\text{LBP}^{\text{u2}}_{n,r}$. For example, 01010 has 2 transitions from 0 to 1 and 2 transitions from 1 to 0 which makes 4 transitions in total. This version then $\text{LBP}^{\text{u2}}_{n,r}$ does not count the number 01010. On the other hand, 00111 has only 1 transition from 0 to 1 so it is accepted.

Gaussian Filtered LBP Finally, Mäenpää and Pietikäinen [5] talk about an extension to the LBP that is also used in the work of Kylberg et al. [4], which consists of sampling neighbours about a central pixel at different radii. Instead of sampling these points equidistantly in a circle, the authors sample according to a Gaussian distribution. Figure 3 shows an example of such a sampling. Kylberg et al. [4] denote it as $\text{LBPF}^{\text{ri}}_{n_1,r_1} + {}^{\text{ri}}_{n_2,r_2} + \cdots + {}^{\text{ri}}_{n_j,r_j}$ for the different numbers of sample points $n_1, n_2, \cdots, n_j$ and different radii distances $r_1, r_2, \cdots, r_j$.

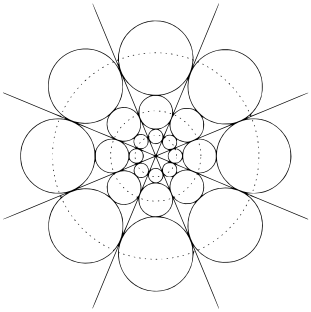


Figure 2: Example of LBPF sampling: In the simple LBP, the points would be sampled at equal distances around the dotted circles for one radius as in figure 1 but with this extension, the points are sampled according to a Gaussian distribution inside the solid black circles at multiple radii. In this picture, the number of neighbours also vary with the radii. This image comes from Mäenpää and Pietikäinen [5].

4.2. Radial Density Profile (RDP)

This method is a way to get the mean intensity in a ring around each pixel in the image. Kylberg et al. [4] defines the *radial mean intensity* f around a center pixel q_c as

$$f(q_c, r) = \frac{1}{|N|} \sum_{q \in N} I(q)$$

where $I(q)$ is the pixel value for pixel q and $N = \left\{q : \|q - q_c\|_2 \in \left(r - \frac{1}{2}, r + \frac{1}{2}\right]\right\}$ is the set of pixels in a ring around q_c of width 1 at distance r from q_c. Figure 3 shows an example of what the set N may look like. Following the notation from Kylberg et al. [4], let $\bar{f}_{q_c}$ be the mean of the set $\{f(q_c, i)\}_{i=1,\cdots,n}$. Then they define the radial density profile for that pixel q_c as $\text{RDP}_n = \left[f(q_c, 1) - \bar{f}_{q_c}, f(q_c, 2) - \bar{f}_{q_c}, \cdots, f(q_c, n) - \bar{f}_{q_c}\right]$.

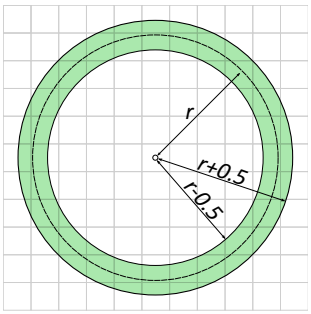


Figure 3: Example of RDP sampling: The green zone is the set of sampled pixels for a given radius r.

It is not explained how the pixel features are combined to make the feature vector for the whole image but we would imagine that the feature vector for the whole image is either a concatenation of the pixel features or the concatenation of the mean of each pixel feature.

Fourier RDP One variation Kylberg et al. [4] introduced is applying the same RDP technique on the image after it undergoes a Fourier transform and looking at it with respect to a base e logarithmic scale. They denote it with FRDP_n.

4.3. Results

The first thing to note is that Kylberg et al. [4] made modifications to the images' scales. In one case, one pixel in the image corresponded to 1 nanometer, they refer to that scale as *fixed scale*. In the other case, the radius of the virus is set to take 20 pixels, they refer to that scale as *object scale*. The classifier used was a Random forest with 100 trees (they tried up to 200 with no significant improvement with respect to 100 trees). The different textures Kylberg et al. [4] used were among

1. $\text{LBP}^{\text{ri}}_{8,2}$
2. $\text{LBP}^{\text{riu2}}_{8,2}$
3. $\text{LBPF}^{\text{ri}}_{8,1} + ^{\text{ri}}_{8,2.4} + ^{\text{ri}}_{8,5.4}$
4. $\text{LBPF}^{\text{riu2}}_{8,1} + ^{\text{riu2}}_{8,2.4} + ^{\text{riu2}}_{8,5.4}$
5. $\text{RDP}_2 0$
6. $\text{FRDP}_2 0$

All values are chosen after a grid search. Figure 4 shows their results with respect to the six aforementioned texture extractors and on both fixed and object scale images. They conclude that $\text{LBPF}^{\text{ri}}_{8,1} + ^{\text{ri}}_{8,2.4} + ^{\text{ri}}_{8,5.4}$ performed the best in fixed scale (median 21% error) whereas $\text{FRDP}_2 0$ performed the best in object scale (median 22% error). They also noticed that combining these two texture extractors, they could get an even better result (median 13% error).

5. Neural Networks Approach

In order to establish a baseline, we combine ideas from Kylberg et al. [4] and ourselves by attempting to use a neural network with the LBP features. Our neural networks are built using the Lasagne and Theano [1, 2] libraries. In our experiments, we will not rescale the images and we will use $\text{LBP}_{8,2}$, an implementation given by the *Mahotas* computer vision library Coelho [3]. We are not scaling the images because we are later comparing to convolutional neural networks which will not used rescaled images. The values of $8, 2$ come from the best values as checked by Kylberg et al. [4]. *Mahotas*'s implementation of LBP is done so that the feature vector is of smaller dimension than the usual dimension of a histogram with 2^N values where N is the number of sampled points. This is achieved by some optimization on their end.

5.1. Results

To normalize the data, we chose to divide the entire dataset by 1.1 times the maximum value of a feature in the training set in order to be more certain that no features in the normalized test set would have a value greater than 1. In all cases, we are using a learning rate of 0.005 and an L2 regularization weight of 0.0001, both arbitrarily chosen and the last layer was a 15 units softmax layer. The dataset described in section 2 is shuffled and the training is done with a batch size of 16 by stochastic gradient descent.

The first feed-forward neural network architecture we used was comprised of 1 hidden layer with 256 units. The results can be found in figures 5 and 6. We note that the best test error was about 52.22%.

The second feed-forward neural network architecture we used was comprised of 2 hidden layers with 256 units each. The results can be found in figures 7 and 8. We note that the best test error was about 52.22%.

The third feed-forward neural network architecture we used was comprised of 3 hidden layers with 256, 128 and 64 units. The results can be found in figures 9 and 10. We note that the best test error was about 51.94%.

5.2. Discussion

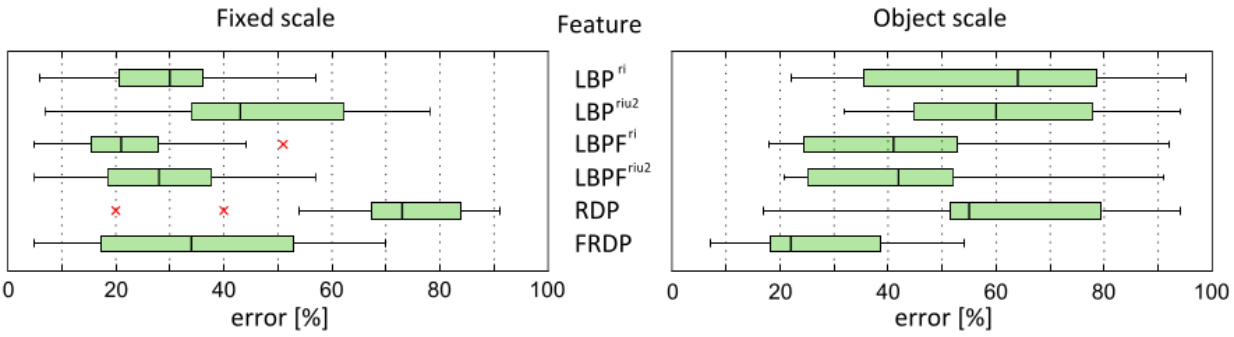
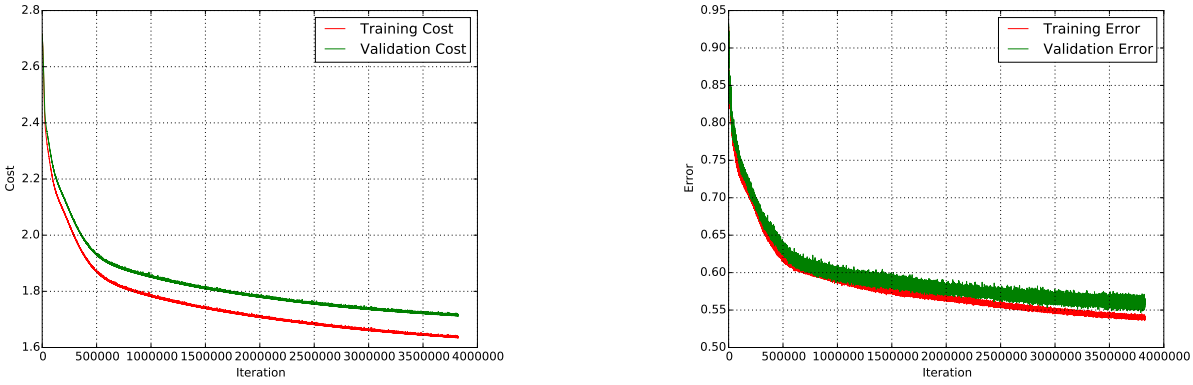


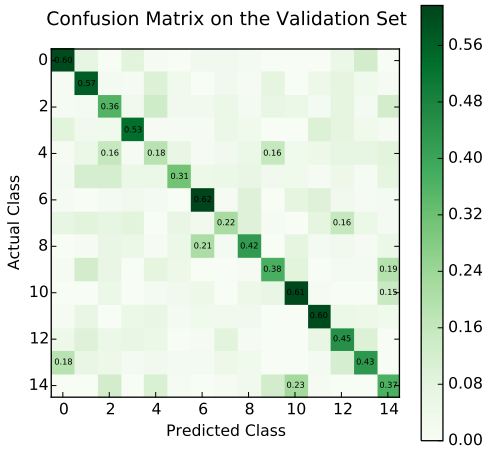
Figure 4: This figure comes from Kylberg et al. [4]. The classification errors are shown for a Random Forest classifier using the 6 different texture extractors as described in section 4.3. The boxes' vertical lines represent the median and the red × represent outliers that are at least 1.5 times the size of the box away from it. The error bars are from the lower to the upper quartile.



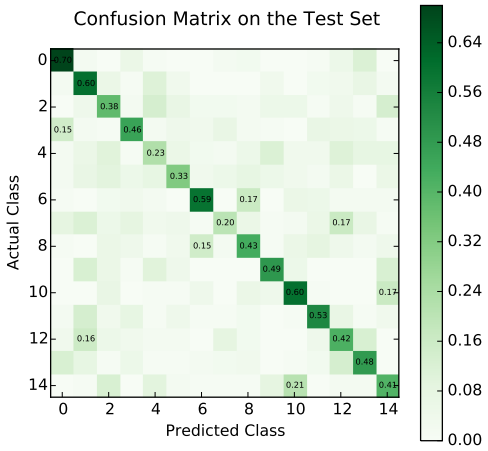
(a) Training versus Validation Cost

(b) Training versus Validation Error

Figure 5: Learning curves for a feed-forward neural network with one hidden layer of 256 units.

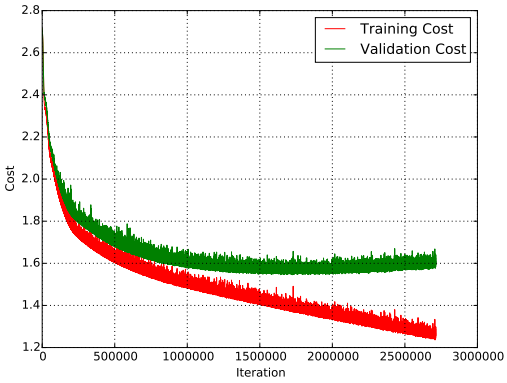


(a) Confusion Matrix on the Validation Set

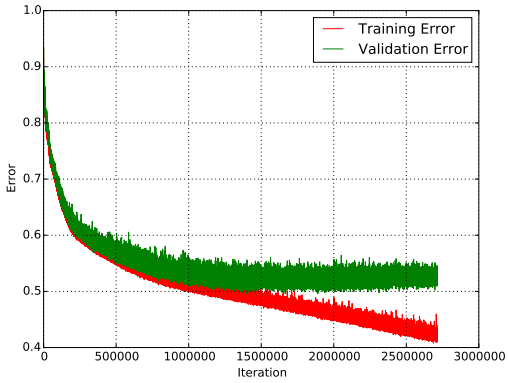


(b) Confusion Matrix on the Test Set

Figure 6: Confusion Matrices for a feed-forward neural network with one hidden layer of 256 unit.

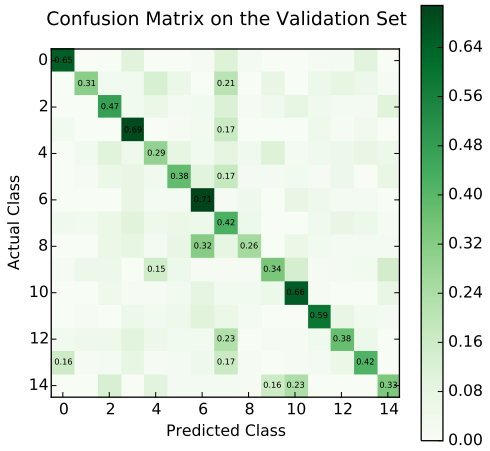


(a) Training versus Validation Cost

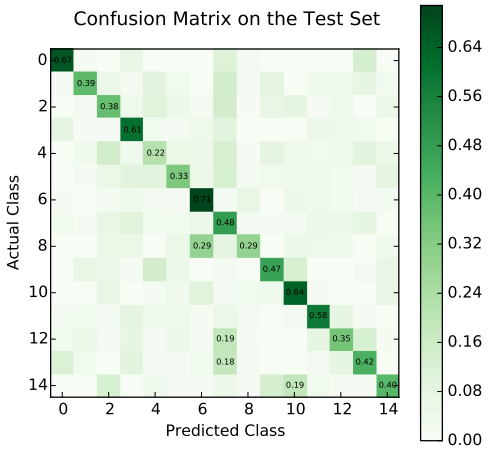


(b) Training versus Validation Error

Figure 7: Learning curves for a feed-forward neural network with two hidden layers of 256 units each.

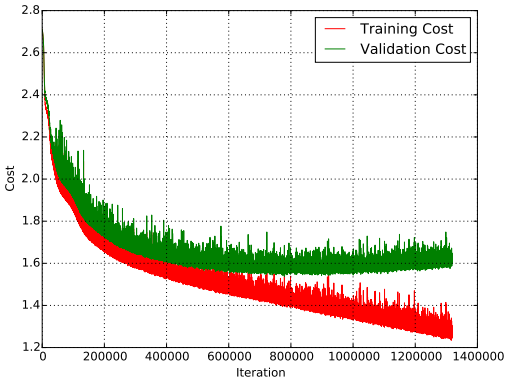


(a) Confusion Matrix on the Validation Set

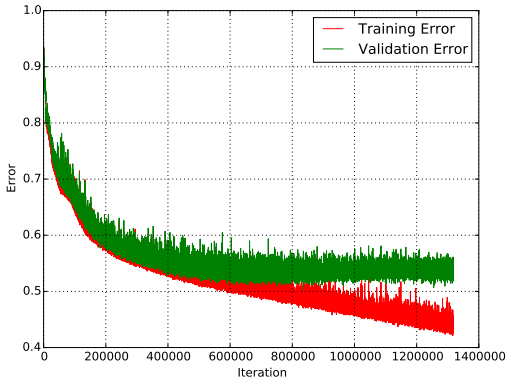


(b) Confusion Matrix on the Test Set

Figure 8: Confusion Matrices for a feed-forward neural network with two hidden layers of 256 units each.

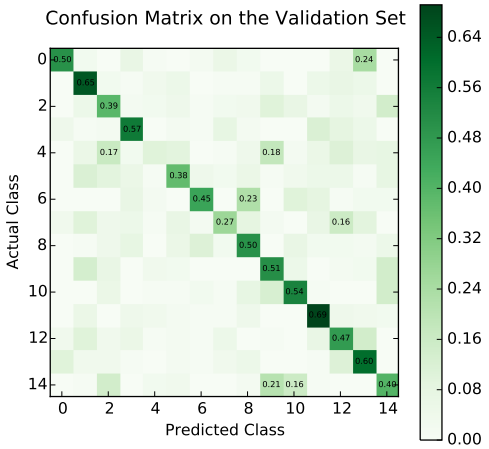


(a) Training versus Validation Cost

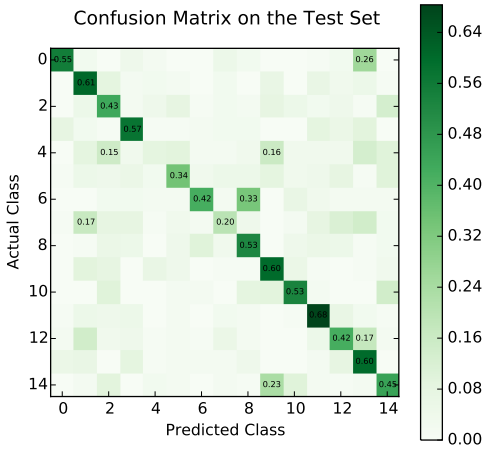


(b) Training versus Validation Error

Figure 9: Learning curves for a feed-forward neural network with three hidden layers of 256, 128, 64 units.



(a) Confusion Matrix on the Validation Set



(b) Confusion Matrix on the Test Set

Figure 10: Confusion Matrices for a feed-forward neural network with three hidden layers of 256, 128, 64 units.

References

[1] Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, James Bergstra, Ian J. Goodfellow, Arnaud Bergeron, Nicolas Bouchard, and Yoshua Bengio. Theano: new features and speed improvements. Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop, 2012.

[2] James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. Theano: a CPU and GPU math expression compiler. In *Proceedings of the Python for Scientific Computing Conference (SciPy)*, June 2010. Oral Presentation.

[3] Luis Pedro Coelho. Mahotas: Open source software for scriptable computer vision. *arXiv preprint arXiv:1211.4907*, 2012.

[4] Gustaf Kylberg, Mats Uppström, and Ida-Maria Sintorn. Virus texture analysis using local binary patterns and radial density profiles. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 573–580. Springer Berlin Heidelberg, 2011.

[5] Topi Mäenpää and Matti Pietikäinen. Multi-scale binary patterns for texture analysis. In *Image Analysis*, pages 885–892. Springer, 2003.