AN EVALUATION OF ROTATION-EQUIVARIANT CONVOLUTIONAL NEURAL NETWORKS

DATA MINING FINAL PROJECT

Gabriel Raya: s1022212, Mikołaj Bocheński: s1042716
Radboud University, The Netherlands

May 9, 2020

ABSTRACT

We evaluate the performance improvement in a medical image classification task which arises from using a rotation-equivariant block in a convolutional neural network, as opposed to using traditional convolutions. We find that performance is not improved much by using rotation equivariance: it yields an AUC score of 0.8949 while traditional convolutions yield 0.8823. Our work, complete with a clean Python implementation of rotation-equivariant networks, is made available online at https://github.com/gabrielraya/Data-Mining/tree/master/Project.

***K*eywords** Convolutional Neural Netowrks · Equivariant CNNs · Histopathology · Medical Imaging

1 Introduction

Computer vision poses many challenges, one of them being the fact that a traditional approach to it - vectorizing images and running a machine learning algorithm on the resulting vectors - has a major pitfall: it is not *translationally equivariant*. To illustrate what this means, let us consider the case of classifying medical images: a tumor is a tumor regardless of whether it appears in the left or the right part of the image, ie. regardless of whether it is *translated* in its entirety. However, after vectorization, these two tumors might look very different to a machine learning algorithm - it does not grasp the inherent structure of images, ie. the importance of the relationships of neighboring pixels.

This problem is solved by using *convolutional neural networks*, which apply trainable filters to each part of the image in separation, and then combine the results. What is important here is the fact that the set of filters is the same for each part of the image, so - going back to our example - the tumor on the left is processed in the same way as the tumor on the right. Only when the image filtering results are combined can differences between the left and the right side of the image become significant - this is, however, desired behavior: one might imagine a classification task in which the left/right position of something in an image matters.

Translational equivariance is not, however, the only desired property of a computer vision algorithm: a slightly smaller tumor is still a tumor, as is a tumor that is rotated 90° clockwise. This creates the need for *scale-* and *rotation-equivariance*, respectively. In this paper, we investigate the effectiveness of rotation-equivariant convolutional neural networks as applied to the problem of histopathologic cancer detection in medical images. Specifically, we compare results obtained using a rotation-equivariant neural network to those obtained using a regular convolutional one.

2 Related work

Both scale- and rotation-equivariance have already been studied; both typically involve applying a convolutional neural network to several copies of the same image, with each copy being either scaled with respect to the original, or rotated. Equivalently, this can be viewed as applying several copies of the same set of filters on the image, with each copy being - again - either scaled or rotated. Our paper is closely connected to [5], in which rotation-equivariant neural networks are applied to medical images. However, in this publication, we focus on comparing different machine learning approaches, whereas the aforementioned paper focuses on applying rotation equivariance to a problem.

3 Mathematical description

3.1 Standard convolutional networks

The network which we base our work on, and which we compare our results to, is a standard convolutional neural network as applied to classification. It consists of two parts: the convolutional part and the fully-connected part. The convolutional part is the first to see input images, it produces outputs which are then flattened and fed into the fully-connected part. This type of architecture is visualized in figure 1.

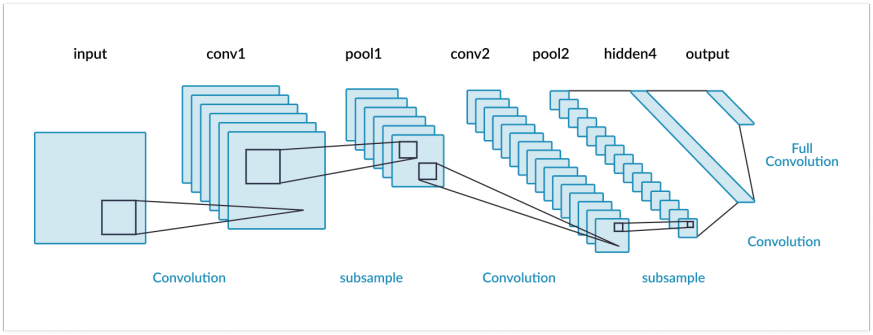

Figure 1: An example standard convolutional architecture.

At each layer l^k of a standard convolutional neural network, a stack of filters ψ is applied to the input feature map f, ie. for layer l^0 this is the input image, and for the other layers this is the previous layer's output. This is formally described in equation 1.

$$[f * \psi^i](x) = \sum_{y \in \mathbf{Z^2}} \sum_{k=1}^{K_f} f_k(y) \psi_k^i(y - x) \tag{1}$$

Here, $*$ denotes the correlation operation[1], K_f denotes the number of channels in the input feature map f, and ψ^i denotes one specific filter. $f_k(y)$ means the pixel value at the k-th channel of image f at position y, similarly for $\psi_k^i(x - y)$.

3.2 Rotation-equivariant networks

The networks whose effectiveness we investigate here are both rotation- and flip-equivariant, meaning that they are equivariant with respect to both rotations of the input image (by a multiple of $90°$) and mirror-flips. This means that there are eight transformations of the image with respect to which we would like processing to be equivariant. Figure 2 shows all of those.

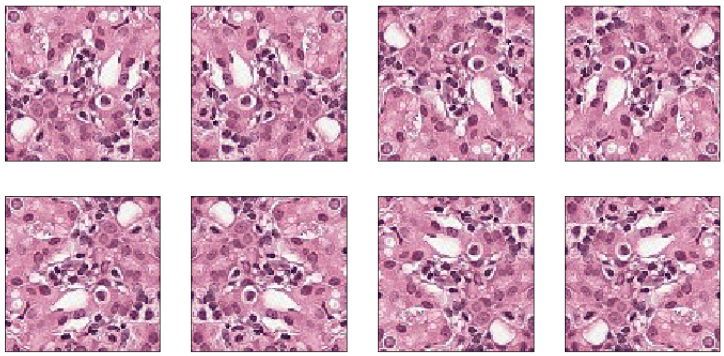

Figure 2: All transformations of example image with respect to which our model is equivariant.

[1] "Convolutional" neural network is a misnomer, CNNs actually use the correlation operation.

Formally, equivariance is a mathematical property of an operation with respect to a transformation, which holds when the operation and transformation can be applied in any order and produce the same result. Equation 2 presents the formalism as described by [2].

$$\Phi(T(x)) = T(\Phi(x)) \tag{2}$$

This simply means that transforming an input x by a tranformation T and then passing it through the layer Φ should give the same result as first mapping x through Φ and then transforming the representation.

One would imagine that to achieve rotation- and flip-equivariance, eight copies of the convolutional part of the network would be applied to the image, with each copy rotated/flipped with respect to the original in a unique way. Then, the outputs of these networks would be pooled (in our case, we use a simple mean over copies) and fed into the fully-connected part of the network, producing a final result.

When implementing the network, we exploit a simple observation: the application of a copy of the convolutional part of our network rotated by eg. $90°$ clockwise to an image is equivalent to the application of a non-rotated copy of the convolutional part to an *image* rotated $90°$ counterclockwise, and then rotating this result back. This is our main innovation; it makes implementing the neural network simpler, but should theoretically give the same results.

3.3 Computational complexity

The computational complexity of the proposed approach is fairly straightforward: each convolutional layer of F filters of size KxL is applied to an image of size NxM with C channels for a complexity of $\mathcal{O}(CFKLMN)$ in both the forward and the backward pass. The number of computations used in a rotation-equivariant layer is 8 times that of an equivalent standard convolutional layer, because each layer can be seen as being applied 8 times.

4 Experiments & results

4.1 The task

We compare the performance of a rotation-equivariant convolutional neural network with a regular convolutional neural network on a Kaggle competition, Histopathologic Cancer Detection[1]. This is a binary classification task in which 96x96 color images are labelled as either “healthy tissue” or “tumor”. It consists of a massive training set of 220025 images, as well as a test set which we do not use due to the unavailability of labels. Instead, we further split the provided training set into an actual training set (the same for both neural networks) and a validation set, on which performance evaluation is carried out.

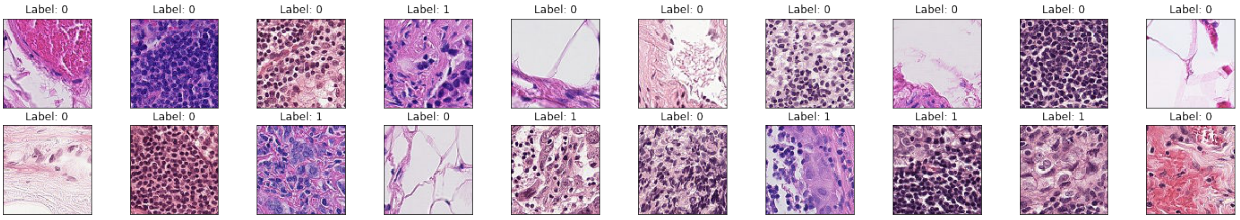

Figure 3: 20 sample images from the training set. Labelling: 0 = healthy, 1 = cancer.

As can be seen in figure 3, it is very difficult to distinguish cancer from healthy tissue just by manual inspection.

4.2 Neural architecture

To rule out any factors that could bias our results, we used the exact same architecture for both the standard and the rotation-equivariant network, with the only difference being the type of the convolutional layers. We have tested several architectures and picked one which gives good overall performance.

1. convolutional: $3 \rightarrow 8$, 3x3

2. 2x2 max-pooling
3. ReLU activation
4. convolutional: $8 \rightarrow 16$, 3x3
5. 2x2 max-pooling
6. ReLU activation
7. convolutional: $16 \rightarrow 32$, 3x3
8. 2x2 max-pooling
9. ReLU activation
10. linear: $512 \rightarrow 256$
11. ReLU activation
12. linear: $256 \rightarrow 128$
13. ReLU activation
14. linear: $128 \rightarrow 2$
15. softmax activation

The architecture is visualized on figure 4. Note that only a 32x32 image is fed into the network - this is due to the fact that the tumor detection is supposed to only be done on the central 32x32 area of each image, with the rest being provided as context.

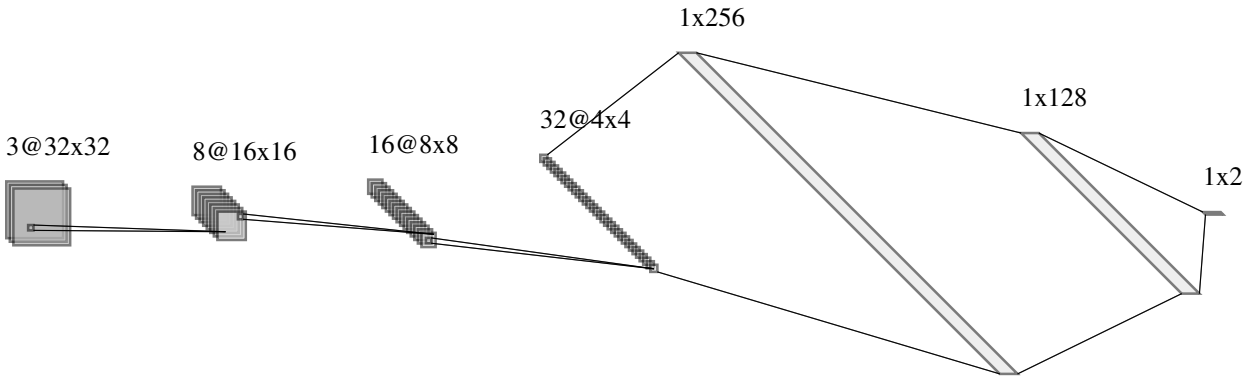

Figure 4: The architecture of our networks. Convolutional layers have been grouped with their pooling and activation.

The other architectures were rather similar, but tended to give worse results. However, they were all characterised by only small differences in the performance of the standard and rotation-equivariant models. They can be viewed in the commit history at https://github.com/gabrielraya/Data-Mining/commits/master.

4.3 Results

Against our expectations, the standard and equivariant models did not differ much in terms of performance by any measure. For example, in the receiver operant characteristic curve comparison in figure 5, we see that the equivariant model performs slightly better between a false positive rate of 0.2 and 0.4, but this difference is only one of a few percentage points. For the model's AUC scores and accuracies, see table 4.3.

Model	**Accurary**	**AUC**
Standard CNN	0.7929	0.8823
Equivariant CNN	0.7968	0.8949

Table 1: AUC and accuracy score comparison

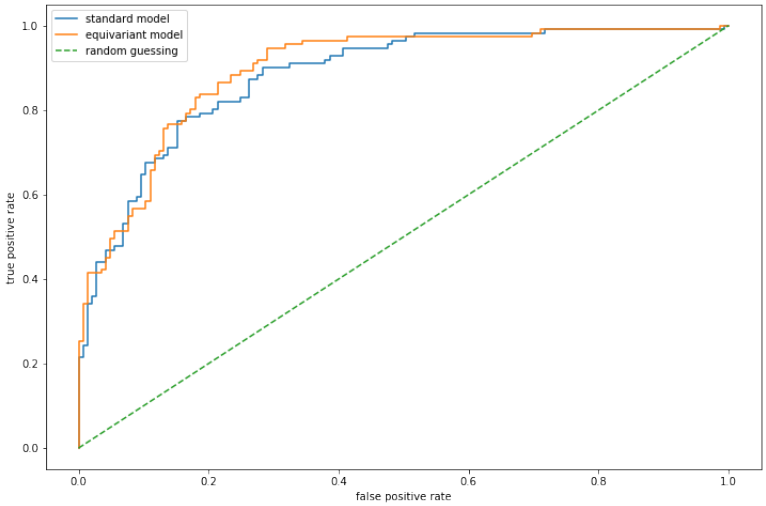

Figure 5: ROC curve comparison

To more objectively evaluate this difference, we perform a sign test ($H0$: the equivariant model performs the same or worse as the standard model, H_A: the equivariant model outperforms the standard model) and obtain an insignificant p-value 0.105.

For completeness, in figure 6, we also present the confusion matrices for the each of the models at a cut-off threshold of 50% confidence for the "tumor" class. As can be seen, these matrices do not differ significantly.

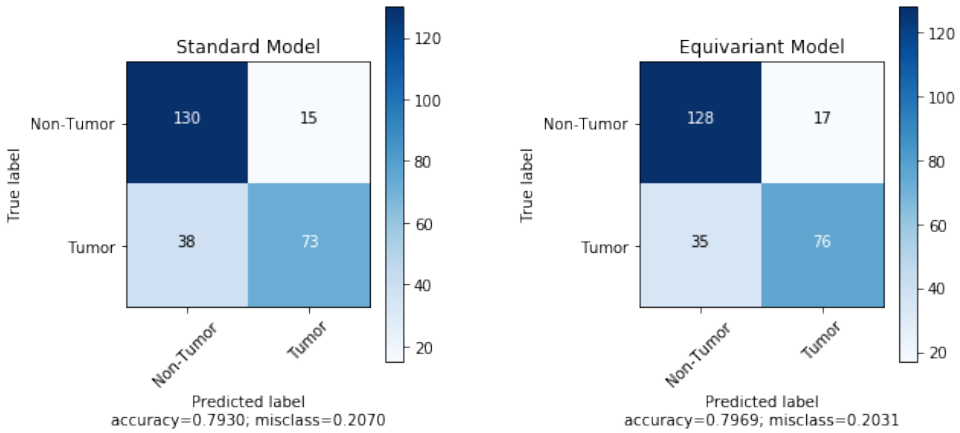

Figure 6: Confusion matrix comparison

To investigate if there are, in fact, any significant differences between the models in terms other than performance, we plot the confidence score for the "tumor" class as given by the equivariant model against that given by the standard model in figure 7. As can be seen, the scores assigned to samples by the models are quite similar, meaning that they are likely paying attention to similar aspects of the image.

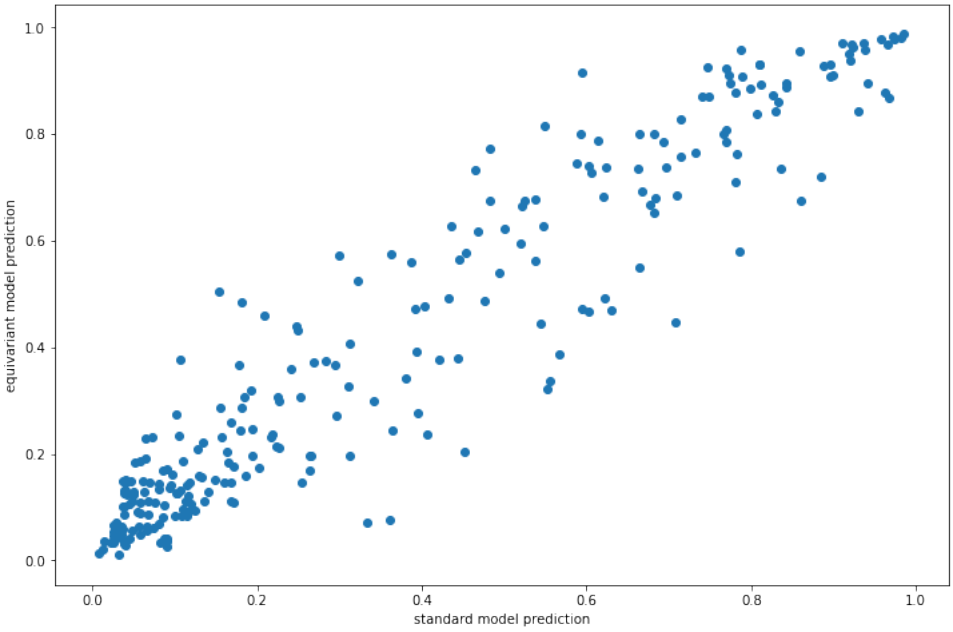

Figure 7: Probability of belonging to "tumor" class, as estimated by different models.

To further investigate the differences between the models, we visualize their internal activations at different layers when viewing an example image - see figure 8. Again, it can be seen that they are processing images in similar ways - they focus on the same groups of cells.

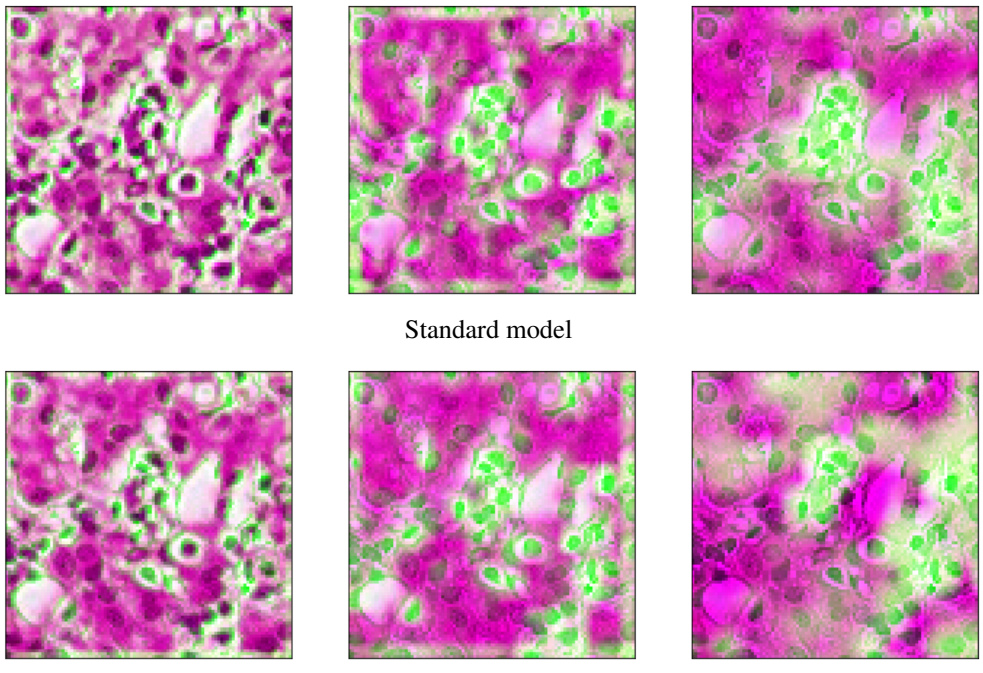

Figure 8: Internal model activations (green, pooled over channels) when looking at example image (pink).

5 Conclusions & future work

On this particular dataset, and for this particular architecture, rotation equivariance proved not to provide a particularly significant boost in network performance. However, to fully evaluate the strengths and weaknesses of the rotation-equivariant layer, one would need to perform an evaluation on many more datasets and using many more architectures. Especially the architectures would need scrutiny; we suspect that the observed lack of difference between the equivariant and standard models is due to the fact that the amount of parameters in the standard model was enough to examine the image under all rotation/flip combinations. This is room for future work.

To facilitate this, we publish our code, which contains a clean Python implementation of rotation-equivariant convolutional neural networks. This implementation utilizes Pytorch[3]; we use Scikit-Learn[4] for evaluation.

References

[1] Kaggle competition: Histopathologic cancer detection. https://www.kaggle.com/c/histopathologic-cancer-detection/.

[2] Taco S. Cohen and Max Welling. Group equivariant convolutional networks. *CoRR*, abs/1602.07576, 2016.

[3] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[5] Bastiaan S. Veeling, Jasper Linmans, Jim Winkens, Taco Cohen, and Max Welling. Rotation equivariant cnns for digital pathology. *CoRR*, abs/1806.03962, 2018.