Rotation invariance for unsupervised cell representation learning

Analysis of the impact of enforcing rotation invariance or equivariance on representation for cell classification

Philipp Gräbel[1], Ina Laube[1], Martina Crysandt[2], Reinhild Herwartz[2], Melanie Baumann[2], Barbara M. Klinkhammer[3], Peter Boor[3], Tim H. Brümmendorf[2], Dorit Merhof[1]

[1]Institute of Imaging and Computer Vision, RWTH Aachen University, Germany
[2]Department of Hematology, Oncology, Hemostaseology and Stem Cell Transplantation, University Hospital RWTH Aachen University, Germany
[3]Institute of Pathology, University Hospital RWTH Aachen University, Germany
graebel@lfb.rwth-aachen.de

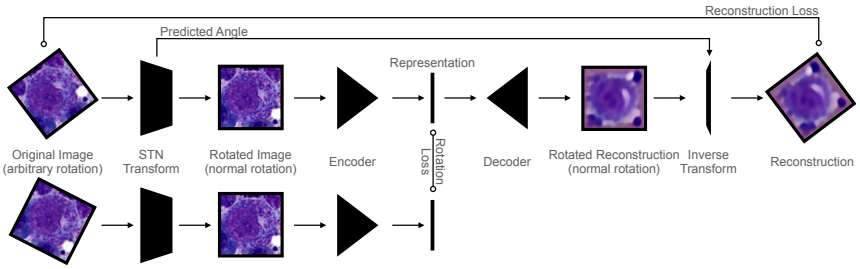
Abstract. While providing powerful solutions for many problems, deep neural networks require large amounts of training data. In medical image computing, this is a severe limitation, as the required expertise makes annotation efforts often infeasible. This also applies to the automated analysis of hematopoietic cells in bone marrow whole slide images. In this work, we propose approaches to restrict a neural network towards learning of rotation invariant or equivariant representation. Even though the proposed methods achieve this goal, it does not increase classification scores on unsupervisedly learned representations.

1 Introduction

Analysis of hematopoietic cells in bone marrow samples is a critical step for diagnosis of many hematological diseases, e.g. leukemia. Currently, medical experts have to manually perform the tedious task of identifying and counting a large number of cells in bone marrow slides. An automated analysis, particularly the classification of various cell types, is a challenging problem but could potentially improve throughput as well as objectivity.

While supervised learning of deep neural networks was shown to be a promising approach [1], it requires a large number of manually created expert annotations. Since this is a time-consuming task, it is infeasible to rely solely on fully supervised methods. Cell detection, however, is comparatively simple [2]. Furthermore, manual validation of automated detection results can be performed by non-experts in short time. Consequently, there is an abundance of patches centered around an individual cell (of unknown type) that can be used for unsupervised representation learning.

Fig. 1. Overall pipeline.



In this work, we focus on auto-encoders [3], which allow extracting a representation of an image in the bottleneck by minimizing a reconstruction loss. As cells occur in arbitrary orientation, a network often learns different representations for the same cell type – even when using rotation augmentations. Finding rotation invariant representations is not straight-forward as the auto-encoder requires orientation information to reconstruct an image. As the classification of cell types is inherently rotation invariant it would be desirable to have rotation invariant representations as well.

A typical solution is data augmentation [4]: training a network with arbitrarily rotated images to force it to learn valid representations for each angle. However, the network still learns multiple representations even for the same image. Intrinsically rotation invariant operations in the network architecture therefore offer a more suitable solution. The HNet [5] uses harmonic convolutions instead of classical convolutional layers to make each operation rotation invariant or equivariant (depending on the chosen rotation order). Even in theory, however, a rotation invariant network cannot be used for reconstruction with the orientation information missing in the representation. Additionally, the proposed HNet is shallow and not suitable for the complex data of hematopoietic cells.

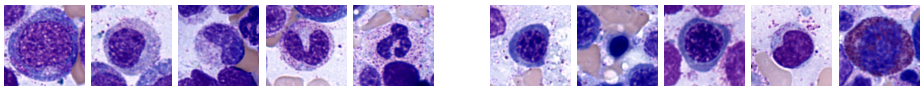
We thus propose the following approach (Fig. 1): A spatial transformer network (STN) [6] is used to normalize the images in rotation direction by minimizing the Kullback Leibler (KL) Divergence [7] between two rotated versions of the same image. A rotation invariant network architecture finds a rotation invariant representation of these normalized images. All methods are further evaluated with respect to the suitability of learned representations for supervised cell classification.

2 Materials and methods

2.1 Image data

The dataset consists of several Whole Slide Images (WSI) of human bone marrow, acquired with a 63× magnifying lens and automated immersion oiling. Each sample is pre-processed with Pappenheim staining to highlight hematologically relevant structures. In this work, we utilize patches of size $256 \times 256\,\mathrm{px}^2$ centered around an individual cell. The cell positions are determined automatically

Fig. 2. One sample patch of each cell type. Left (neutrophilic granulocytes): promyelocyte, myelocyte, metamyelocyte, band and segmented granulocyte. Right: polychromatic, orthochromatic and basophilic erythroblast, lymphocyte and eosinophilic granulocyte.



using U-Net [8] and *Watershed* [9] and subsequently manually validated and, if necessary, corrected. Hematological experts assigned each patch a cell type corresponding to the cell in the center, which might be surrounded by other cells as well. This results in a dataset of 6 085 samples of ten different cell types, as shown in Fig. 2.

A larger unlabelled dataset is employed to learn representations in an unsupervised fashion. This dataset contains approx. 11 000 patches centered around a cell of unknown cell type each from different images than the fully labelled dataset. It should further be noted that the influence of surrounding cells on image reconstruction tasks is reduced through decreasing the influence of pixels on the loss based on their distance to the center. This is a necessary prerequisite as surrounding cells are neither generally relevant for the classification of the cell nor rotation-invariant per se.

2.2 Rotation Loss

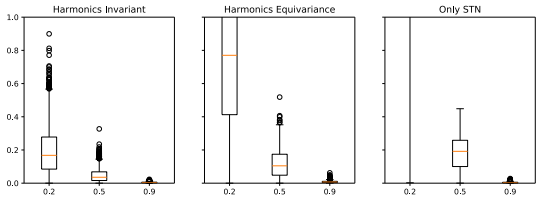
We apply a rotation loss by sending two arbitrarily rotated versions of an image through the network and comparing the output representations. As the representation of a Variational Auto Encoder [10] is described by a Gaussian distribution with mean μ and variance σ^2, we can compute the similarity between representations using the Kullback Leibler (KL) divergence [7] d_{KL}. We use the symmetric KL divergence of the two representations as an additional loss term, which penalizes representations that differ from the representations of rotated versions of the same image

2.3 Harmonic Convolutions

Using spherical harmonics in convolutional layers, a specific rotation order can be enforced as shown in the *HNet* [5]. In our work, we use the proposed definitions for invariance and equivariance: a mapping is rotation equivariant if and only if each rotation in the image domain can be associated with a unique transformation in the feature domain. For invariance, the transformation needs to be the identity function.

In HNet, rotation equivariance is achieved by restricting the filters of convolutional layers to be of the form $W_m(r, \theta, R, \beta) = R(r)e^{i(m\theta+\beta)}$, where (r, θ) are the polar coordinates of the input feature map, R and β are learned by the network and the rotation order r is a meta parameter that is chosen within the architecture design. Rotational invariance requires rotation order $r = 0$, while

Fig. 3. Results in terms of KL Divergence between the representations of different rotations of test images. For each of the networks types, three different rotation loss weight factors are tested.



equivariance only requires the existence of a unique rotation order. In this paper, rotation order $r = 2$ is chosen for rotation equivariant networks.

Since the originally proposed network is comparatively shallow, we introduce a ResNet-like architecture that is deeper and has residual connections [11]. In order to keep the rotation order consistent, we enforce that the rotation order of sub-networks skipped by a residual connection is zero. We further apply Harmonic Batch Normalization. As it is not possible to use a stride larger than one or maximum pooling with equivariant convolutions, we employ average pooling for downsampling.

2.4 Experimental setup

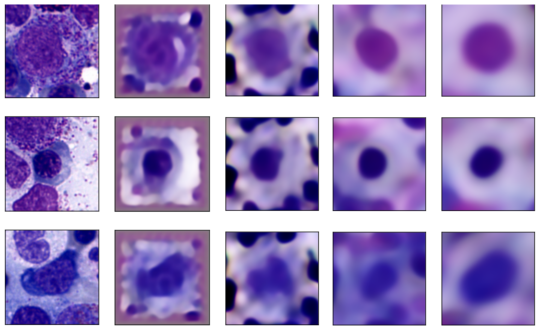
For the evaluation, several VAEs are trained unsupervisedly using reconstruction losses to learn a useful representation of hematopoietic cells. To this end, the dataset with unlabeled cell patches is employed in six-fold cross-validation. Training is performed using the Adam optimizer until an early stop criterion based on a separated validation set is reached. First of all, the reconstruction quality is evaluated in terms of SSIM between original and reconstructed images of the test sets. Secondly, the network is used to extract representations (feature vectors) of the dataset with labelled cell patches. These vectors are used as input to a shallow classification network, which is evaluated in terms of F1-score in five-fold cross-validation.

Each VAE has a Spatial Transform Subnetwork and utilizes the weight rotation loss wL_{rot} with the weight $w \in [0.2, 0.5, 0.9]$. The network is either a classical VAE (denoted as *only-STN*), a VAE with harmonic layers of rotation order zero (*inv-HNet*) or a VAE with harmonic layers of rotation order two (*equ-HNET*). As a baseline, we us a normal VAE without additional efforts to establish rotation invariant representations.

3 Results

Fig. 3 shows the KL Divergence between representations of test images in different rotations. Fig. 4 shows reconstruction results for three sample images. All methods yield representations that fulfill the desired condition: being invariant with respect to rotation of input images. As expected, this effect is stronger

Fig. 4. From left to right: original image, baseline reconstruction, reconstructions from only STN, equivariant HNet and invariant HNet (with rotation loss weight $w = 0.5$ each). Note that effects at the patch border are due to the surrounding cell suppression.



with larger weights for the rotation loss. Visual inspection of the reconstructions suggests that the network achieves rotation invariance or equivariance mostly through reconstructing a rotation symmetric image.

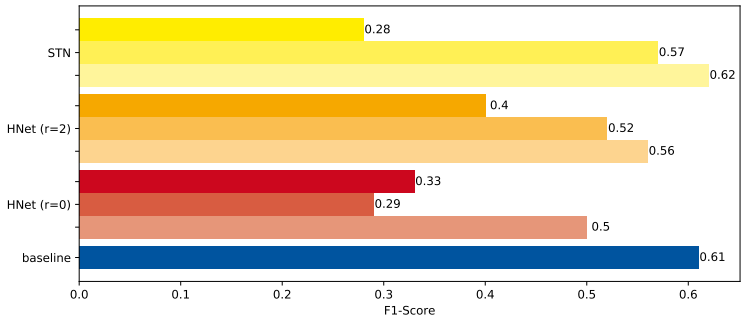
Fig. 5 shows the corresponding classification results on labelled cell images. It can be seen that the representations from networks trained with rotation invariance methods have generally lower or similar scores compared to the baseline. For most methods, a higher rotation loss weight yields a lower classification score.

4 Discussion

The qualitative results suggest that the presented methods to obtain rotation invariant or equivariant representations mostly achieve this by learning rotation symmetric reconstructions. While this lowers the reconstruction accuracy slightly, it has a larger impact on classification accuracy with learned representations. Enforcing rotation invariant harmonic convolutions (rotation order zero) is most detrimental to the F1-score, while harmonic networks with rotation order two perform slightly better. The trend generally shows that the learned embeddings are less descriptive with respect to the classification of cell types with higher focus on rotation invariance or equivariance. With low rotation loss weight and no harmonic convolutions, similar results compared to the baseline can be reached while having a more rotation independent model.

It remains to be evaluated whether the restriction to a rotation invariant representation is a beneficial constraint in semi-supervised learning approaches. As it has been shown that domain-dependent suitable constraints improve semi-supervised strategies, it could be a valuable approach in these settings. Furthermore, losses that penalize purely rotation symmetric reconstructions for non-symmetric images might increase the usefulness of the presented methods. Further research should include the amount of labelled data as well as additional augmentations.

Fig. 5. Classification results using the learned representations. Lighter colors indicate lower values for the rotation loss weight (from top to bottom: 0.2, 0.5 and 0.9).



Acknowledgement. This study was supported by the following grants: DFG: SFB/TRR57, SFB/TRR219, BO3755/6-1, STE 2802/1-1, BMBF: STOP-FSGS-01GM1901A, BMWi: EMPAIA project to PB.

References

1. Gräbel P, Crysandt M, Herwartz R, et al. Evaluating Out-of-the-box Methods for the Classification of Hematopoietic Cells in Images of Stained Bone Marrow. 1st MICCAI Workshop on Computational Pathology (COMPAY). 2018;.
2. Gräbel P, Özcan Özkan, Crysandt M, et al. Circular Anchors for the Detection of Hematopoietic Cells using RetinaNet. IEEE International Symposium on Biomedical Imaging (ISBI). 2020;.
3. Kramer MA. Nonlinear principal component analysis using autoassociative neural networks. AIChE Journal. 1991 02;37:233–243.
4. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25. 2012; p. 1097–1105. Available from: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.
5. Worrall DE, Garbin SJ, Turmukhambetov D, et al. Harmonic Networks: Deep Translation and Rotation Equivariance. CoRR. 2016;abs/1612.04642. Available from: http://arxiv.org/abs/1612.04642.
6. Jaderberg M, Simonyan K, Zisserman A, et al. Spatial Transformer Networks. Advances in Neural Information Processing Systems 28. 2015; p. 2017–2025. Available from: http://papers.nips.cc/paper/5854-spatial-transformer-networks.pdf.
7. Kullback S, Leibler RA. On Information and Sufficiency. Ann Math Statist. 1951 03;22(1):79–86. Available from: https://doi.org/10.1214/aoms/1177729694.
8. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. CoRR. 2015;abs/1505.04597. Available from: http://arxiv.org/abs/1505.04597.
9. Beucher S, Meyer F. The morphological approach to segmentation: the watershed transformation. Optical Engineering-New York-Marcel Dekker Incorporated-. 1992;34:433–433.
10. Kingma D, Welling M. Auto-Encoding Variational Bayes. ICLR. 2013 12;.
11. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. CoRR. 2015;abs/1512.03385. Available from: http://arxiv.org/abs/1512.03385.