Improved Semantic Segmentation for Histopathology using Rotation Equivariant Convolutional Networks

Jim Winkens *
University of Amsterdam
jim.winkens@gmail.com

Jasper Linmans *
University of Amsterdam
jasper.linmans@gmail.com

Bastiaan S. Veeling
University of Amsterdam
bas.veeling@gmail.com

Taco S. Cohen
University of Amsterdam
taco.cohen@gmail.com

Max Welling
University of Amsterdam
welling.max@gmail.com

Abstract

We propose a semantic segmentation model for histopathology that exploits rotation and reflection symmetries inherent in histopathology images. We demonstrate significant performance gains due to increased weight sharing, as well as improvements in predictive stability. The group-equivariant CNN framework is extended for segmentation by introducing a new $(G \rightarrow \mathbb{Z}^2)$-convolution that transforms feature maps on a group to planar feature maps. Also, equivariant transposed convolution is formulated for up-sampling in an encoder-decoder network. We further show the importance of exploiting more symmetries by varying the size of the group.

1 Introduction

Recent advancements in the digitization of microscopic images have motivated the development of image analysis algorithms to assist or automate diagnostic tasks in the field of digital pathology. Here convolutional neural networks (CNNs) have shown their potential for pathologist-level performance in a variety of tasks [1].

A core property that has contributed to the effectiveness of these models is the efficient sharing of parameters in the convolutional layer which induces translation equivariance: shifting a layer's input produces a proportionate shift in the layer's output. In other words, the translation symmetry of images is preserved. However, such layers conventionally only exploit translation symmetries and not the rotation and reflection symmetries that are inherent in whole-slide images (WSIs).

We propose a pixel-wise segmentation model that encodes these symmetries by using group convolutions [2] that have been shown to improve classification performance. We extend this framework with a convolution operation that maps from a group representation to the $\mathbb{Z}^2$ grid and we adopt group convolution in a transposed convolution layer, which we show to be equivariant. We evaluate the model on a Camelyon16 [3] derived dataset, showing that the increased weight sharing by explicitly encoding rotation and reflection symmetries leads to consistent performance gains.

Furthermore, we establish that conventional CNNs trained on histopathology data demonstrate erratic behavior under rotation and reflection. To accommodate the requirements of model predictability in a clinical setting, it is critical to have stability guarantees. We show that the proposed group-equivariant model improves upon the stability under $\pi/2$ rotations and reflections compared to an equivalent standard CNN.

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

*These authors contributed equally to this work

2 Methodology

2.1 Mathematical framework

To exploit rotation and reflection symmetries in a semantic segmentation setting, we extend[1] the mathematical framework of Group equivariant Convolutional Neural Networks (G-CNNs) introduced in [2]. G-CNNs utilize group convolutions, which enjoy increased weight sharing and better statistical efficiency than regular convolutions. Specifically, it is implemented for the $p4$ group, consisting of translations and rotations by multiples of $\pi/2$, and the $p4m$ group, which additionally includes reflections. In this framework, feature maps are considered as functions on these groups, e.g. for the $p4m$ group the orientation channels come in groups of 8 corresponding to the number of roto-reflections in the group. The first layer transforms an input image $f : \mathbb{Z}^2 \to \mathbb{R}^K$, with K the number of channels using a filter ψ, which is defined as the $(\mathbb{Z}^2 \to G)$ convolution:

$$[f * \psi](g) = \sum_{y \in \mathbb{Z}^2} \sum_{k=1}^{K} f_k(y) \psi_k(g^{-1}y), \tag{1}$$

where $g = (r, t)$ is a roto-translation (in case $G = p4$) or a roto-reflection-translation (in case $G = p4m$). In the next layers, the feature maps and filters are both functions on G, for which the $(G \to G)$-convolution is used:

$$[f * \psi](g) = \sum_{h \in G} \sum_{k=1}^{K} f_k(h) \psi_k(g^{-1}h). \tag{2}$$

Note that the transpose of this linear operation, the *transposed* convolution, is also equivariant. Contrary to strided group convolution [2], the choice of stride for transposed group convolution does not affect equivariance.

To allow for the equivariant transformation of a feature map in G to a two-dimensional segmentation mask $m : \mathbb{Z}^2 \to \mathbb{R}^C$, with C the number of classes, we present the $(G \to \mathbb{Z}^2)$-convolution:

$$[f * \psi](y) = \sum_{h \in G} \sum_{k=1}^{K} f_k(h) \psi_k(z(y)^{-1}h), \tag{3}$$

where $z(y)$ is the translation by y in $\mathbb{Z}^2$. This convolution uses a single planar filter per feature map that is shared across the orientation channels to preserve equivariance.

2.2 GU-Net architecture

To obtain pixel-wise segmentation maps, we use the conventional U-Net architecture [4] as a baseline for our rotation equivariant model. The *GU-Net* architecture is constructed by replacing all (transposed) convolution and batch normalization layers with their group equivariant counterparts. Two-layer convolution blocks are followed by a 3×3 max-pool that incrementally reduces the spatial size, up to a depth level of four. Then pooling is replaced by 3×3 transposed convolutions with zero padding to recover the spatial size of the input image and enable per-pixel classification. The proposed $(G \to \mathbb{Z}^2)$-convolution is finally used to transform the orientation channels to the two-dimensional grid of the output mask.

3 Experiments

3.1 Dataset

The proposed model is evaluated on the PatchCamelyon dataset [5] derived from the Camelyon16 challenge [3] with the task of tumor localization. The original challenge data contains 400 H&E stained WSIs of sentinel lymph node sections with pixel-level annotations. The PatchCamelyon dataset consists of 327.680 patches of 320×320 pixels at $10\times$ magnification, with a 8:1:1 split into training, validation and test sets. The patches were extracted by uniformly sampling WSIs and drawing tumor/non-tumor patches with equal probability.

[1] Implementations of the equivariant layers are available at `https://github.com/JasperLinmans/gcnn-segmentation`.

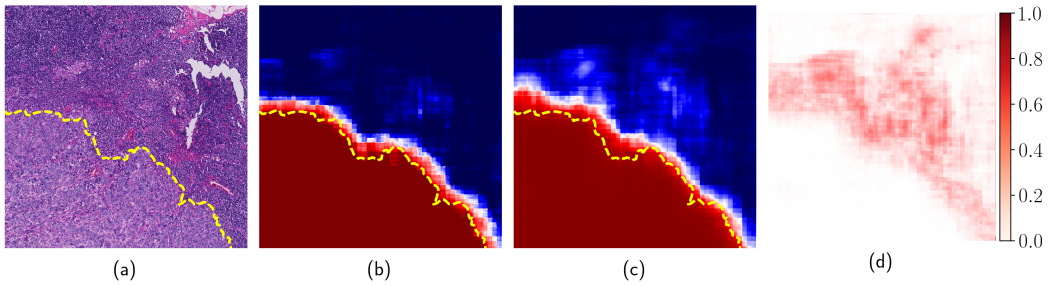

Figure 1: **(a)** Example of a $630\,\mu\mathrm{m}^2$ patch and its ground truth tumor mask (marked by the yellow border), **(b)** the mean tumor prediction of the patch across the eight roto-reflection transformations in $p4m$ as viewed in the original orientation using the P4M U-Net and **(c)** using the baseline U-Net. **(d)** Standard deviation of the tumor predictions across the roto-reflection transformations using the baseline U-Net. The standard deviation of the P4M U-Net model is omitted since the values are not visible in the same range, as all values are lower than $1\mathrm{e}{-6}$.

3.2 Model stability

To assess the predictive stability of the P4M U-Net model, we examine the standard deviation of predictive probabilities under roto-reflection transformations of the input as compared to the baseline model. To ensure a fair comparison, the number of parameters is kept constant [2] by dividing the number of filters per layer by $\sqrt{8} \approx 3$, the square root of the group size. Figure 1 shows the analysis for an example patch. We observe that the standard U-Net is prone to unstable prediction behavior under roto-reflections, especially at the tumor boundaries.

3.3 Segmentation performance

The segmentation performance of the proposed model is evaluated using the Dice similarity coefficient. To compare using group convolutions to a data augmentation strategy, the dataset was augmented with roto-reflection transformations for the baseline model. We report Dice coefficients on the test set of 0.80, 0.83 and 0.84 for the baseline U-Net, the P4 U-Net and the P4M U-Net respectively. These results show that explicitly encoding rotation and reflection symmetries leads to consistent performance gains. We further observe better performance when exploiting more symmetries due to increased weight sharing.

4 Conclusion

By extending the group-equivariant CNN framework and using it in an encoder-decoder architecture for pixel-wise predictions, we show a consistent performance increase compared to an equivalent conventional CNN. Notably, we also demonstrate that using group convolutions results in significantly more stable segmentation, accommodating the model predictability requirements in a clinical setting.

References

[1] Geert Litjens et al. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[2] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning*, pages 2990–2999, 2016.

[3] Babak Ehteshami Bejnordi et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22):2199–2210, 2017.

[4] Olaf Ronneberger et al. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, 2015.

[5] Bas Veeling et al. Rotation equivariant cnns for digital pathology. In *International Conference on Medical image computing and computer-assisted intervention*, 2018.