University of Southampton

Rotation-Scale Equivariant Steerable Filters

Yilong Yang Srinandan Dasmahapatra Sasan Mahmoodi

School of Electronics and Computer Science, University of Southampton

Introduction

Digital histology imaging of biopsy tissue can be captured at arbitrary orientations and magnification, resulting in cells appearing in different scales. Incorporating rotation or scale equivariance into CNNs has proved to be effective in improving models' generalization performance. In this paper, we introduce Rotation-Scale Equivariant Steerable Filter (RSESF), which utilizes filter steerability and Gaussian scale-space theory to parameterize convolutional filters, resulting in an equivariant layer that is stable to rotation and scale variations.

Method

Filter Construction. (The visual illustration is shown in Fig 1)

$$F_k^l(c^l, x, y, \sigma_k^l, \theta_r; c^{l-1}) = \sum_{i,j>0}^{i+j\le N} \alpha^l_{i,j,c^l,c^{l-1}} G^{i,j}_{\theta_r}(x, y; \sigma_k^l), 1 \le k \le \gamma, \theta_r = \frac{2\pi r}{R}, 1 \le r \le R \quad (1)$$

Equivariant Convolution.

The first layer (l=1):

$$f^1_{k,c^1,r}(x,y) = \sum_{c^0=r,g,b} \sum_{x_0,y_0} F_k^1(c^1, x-x_0, y-y_0, \sigma_k^1, \theta_r; c^0)\, f^0_{c^0}(x_0,y_0), 1 \le c^1 \le C^1 \quad (2)$$

Subsequent Layers (l>1):

$$f^l_{k,c,r''} = \sum_{d=1}^{C^{l-1}} \sum_{r'=1}^{R} F^l_{k,c,d,r'} * f^{l-1}_{k,d,r''}, c \in \{1,\cdots,C^l\}, k \in \{1,\cdots,\gamma\} \quad (3)$$

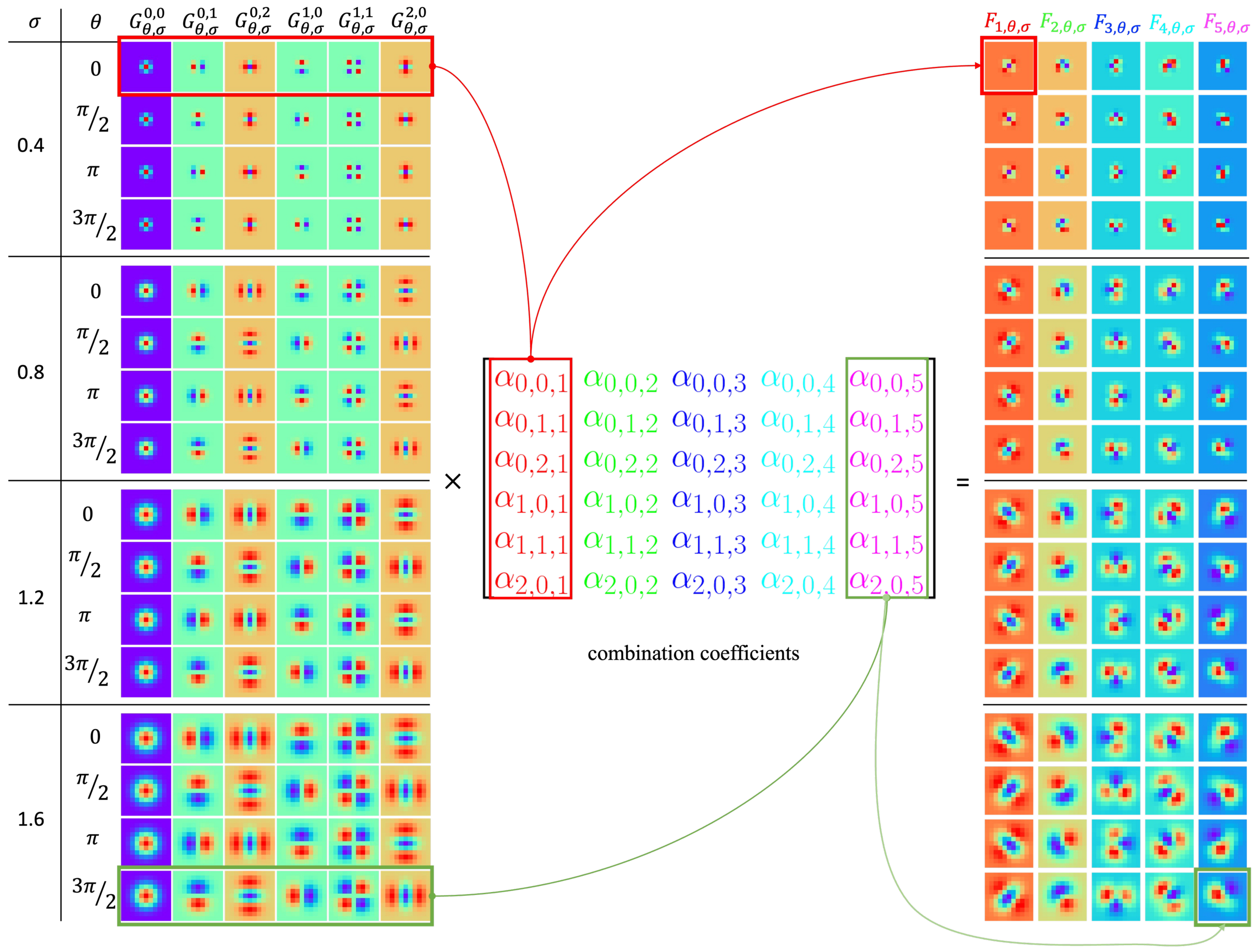

Fig 1. The construction process of an RSESF filter (with 4 scale channels, 4 rotation channels, 1 input channel and 5 output channels) as matrix multiplication.

Memory Efficient Training. Since there is no inter-rotation interaction between rotation channels, we are allowed to train the network within only one rotation channel, but then other R-1 rotation channels are created at inference.

Experiments

Datasets. 1) Gland Segmentation (GlaS) datatset; 2) Colorectal Adenocarcinoma Gland (CRAG) dataset, 3) synthetic texture dataset.

Evaluation Regime. In order to evaluate models' generalization capacity to rotation and scale variations, we design the Out-Of-Distribution (OOD) test, as the complement of the normally used In-Distribution (ID) test. See Table 1 for detail.

Eval Regime	Training Images	Testing Images
ID	Randomly rotated, re-scaled	Randomly rotated, re-scaled
OOD	Original	Randomly rotated, Re-scaled

Table 1. Details of ID and OOD setting.

Results

ID testing. The model with RSESF filters outperforms standard CNN. Interestingly, the RSESF model has significantly fewer parameters, comprising only 4.21% of the CNN model.

OOD testing. RSESF outperforms other compared methods on CRAG and GlaS datasets, achieving competitive results on the texture dataset. Notably, the RSESF achieves state-of-the-art mIoU but is much smaller and more GPU efficient. Table 2 summarizes the comparison between models in terms of segmentation performance, model size and the amount of GPU memory required for training. Fig 3 and Fig 4 show examples of predictions generated by different methods.

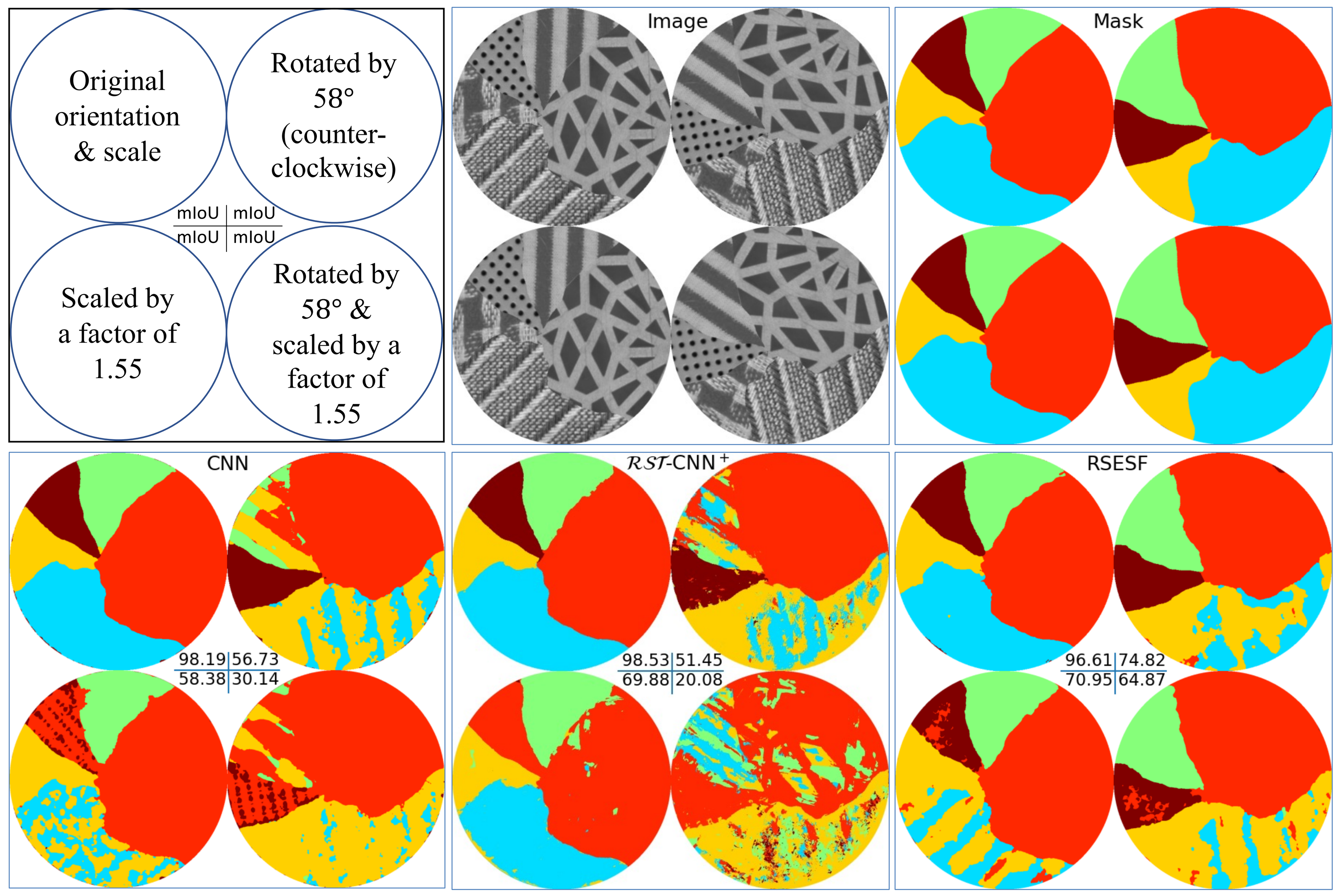

Fig 3. Some visual clues of how equivariant models demonstrate superiority on different versions of test images.

Filter Type	Params (M)	GPU (GB)	N_f	γ	R	ID Testing GlaS*	ID Testing CRAG*	OOD Testing GlaS	OOD Testing CRAG
CNN	29.85	4.48	64	1	1	75.65	75.20	54.19	69.86
RDCF	2.42	5.07	8	1	8	84.38	86.08	55.33	63.38
E(2)CNN	7.03	7.65	8	1	8	**84.19**	**87.91**	52.92	80.79
H-Nets	13.98	7.65	64	1	1	72.47	74.44	57.63	59.34
SDCF	9.66	5.09	16	4	1	83.43	82.75	40.13	68.43
SESN	9.66	5.06	16	4	1	81.77	83.85	51.00	67.51
SEUNet	1.22	5.05	16	4	1	82.50	84.38	58.91	80.30
$\mathcal{RST}$-CNN	0.15	5.06	2	4	8	77.76	78.37	42.90	61.15
$\mathcal{RST}$-CNN$^+$	1.36	15.12	6	4	8	84.08	85.09	56.43	72.02
RSESF	1.22	5.05	16	4	1, 8	83.10	85.00	**60.60**	**82.15**

Table 2. Model Size vs. mIoU vs. GPU Requirement. Columns N_f, γ, R denote the number of filters, scale channels, and rotation channels in the first layer of the UNet, respectively.

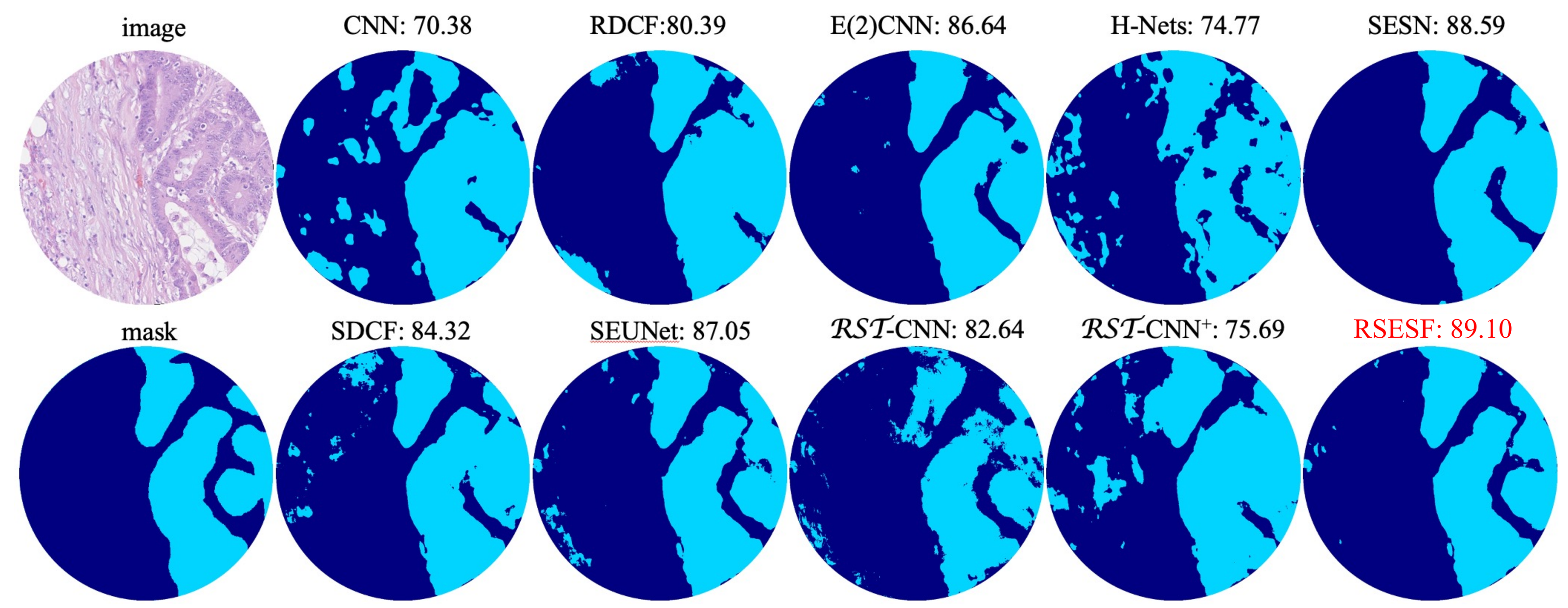

Fig 4. Prediction visualization. mIoU score shown on top of each segmentation map. Image source: CRAG dataset. Light blue: gland; dark blue: non-gland.

Conclusion

We demonstrate that RSESF can generalize CNNs to segment images presented in scales and orientations that do not exist in training samples. Models with RSESF filters have much fewer trainable parameters and can be trained in a memory-efficient way, as the nature of decoupled equivariant convolution gives the model flexibility of training on one orientation but inference in multiple orientations.

Acknowledgement

The authors acknowledge the use of the IRIDIS High-Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work. Yilong Yang is supported by China Scholarship Council under Grant No. 201906310150.

  Contact: Yilong.Yang@soton.ac.uk