IMAGE AUGMENTATION USING RADIAL TRANSFORM FOR TRAINING DEEP NEURAL NETWORKS

Hojjat Salehinejad[*†], *Shahrokh Valaee*[*], *Senior Member, IEEE, Tim Dowdell*[†], *and Joseph Barfett*[†]

[*]Department of Electrical & Computer Engineering, University of Toronto, Toronto, Canada
[†]Department of Medical Imaging, St. Michael's Hospital, University of Toronto, Toronto, Canada
salehinejadh@smh.ca, valaee@ece.utoronto.ca, {dowdellt,barfettj}@smh.ca

ABSTRACT

Deep learning models have a large number of free parameters that must be estimated by efficient training of the models on a large number of training data samples to increase their generalization performance. In real-world applications, the data available to train these networks is often limited or imbalanced. We propose a sampling method based on the radial transform in a polar coordinate system for image augmentation to facilitate the training of deep learning models from limited source data. This pixel-wise transform provides representations of the original image in the polar coordinate system by generating a new image from each pixel. This technique can generate radial transformed images up to the number of pixels in the original image to increase the diversity of poorly represented image classes. Our experiments show improved generalization performance in training deep convolutional neural networks with radial transformed images.

Index Terms— Augmentation, deep learning, imbalanced dataset, polar coordinate system, radial transform.

1. INTRODUCTION

The need for massive amounts of data to train deep neural networks is a major drawback to these models [1], [2]. Generalization performance and versatility of deep learning [3] models are highly dependent on availability of abundant data. The generalization performance refers to the accuracy of the neural network in classification of unseen data. The other challenge is imbalanced datasets, where very few data samples are available for some data classes. These challenges arise in many practical machine learning scenarios such as fraud detection in banking transactions or in medical sciences. In the former, a small number of fraudulent transactions are imbalanced by a high percentage of normal transactions. In the latter, the majority of the population being healthy or a low prevalence rate for certain medical conditions in the dataset can bias the deep learning model.

The training of a neural network with limited data may be mitigated by sampling noise, which exists in the training data but not in the test data drawn from the same distribution [4]. An elegant solution to these challenges is data augmentation, i.e., the application of one or more deformations to a collection of annotated training samples which result in new, additional, and potentially non-redundant training data [5], [6].
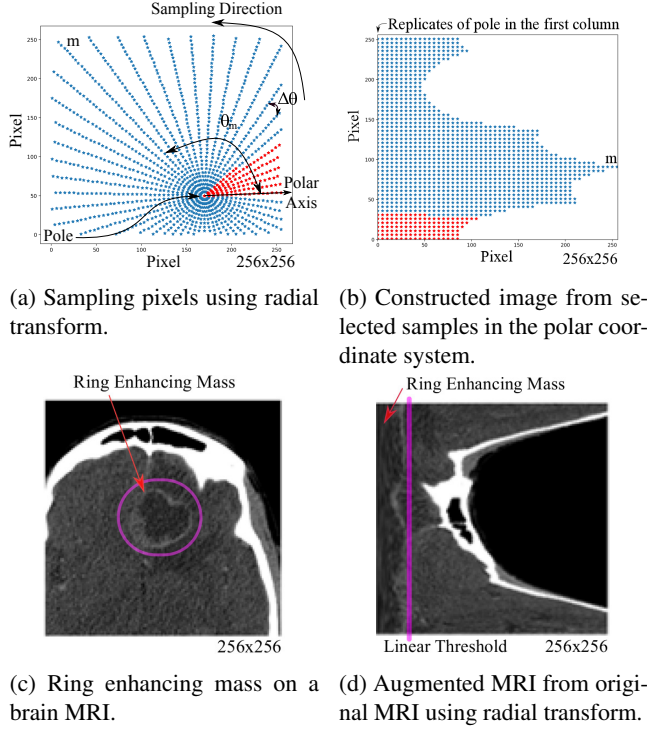
In general, data augmentation does not increase the information content of the dataset. However, it can improve diversity of the dataset and generalization performance. Diversifying the data helps the network to generalize better to unseen data and become invariant to applied deformations [6]. The neural network learns from the added diversity and gains experience in how data belonging to the given labels can "look different". The various deformations commonly applied to labeled data, such as multiplication by a transform matrix, does not affect the semantic meaning of the labels [6]. Some of the image augmentation techniques include adding noise, rotating, translating, mirroring, or scaling the image. Affine is a 2D geometric transform method based on applying a combination of translation, rotation, scaling, and shearing transformations. The Affine augmentation method is widely used as an image augmentation method for correcting geometric distortion introduced by perspective irregularities [7]. Another approach is making an image that contains multiple copies of the original image rotated by different angles [8]. The polar harmonic transform, based on a set of orthogonal projections, is another method that has been used to generate a set of features that are insensitive to rotation [9].

In this paper, an augmentation method based on the radial transform is proposed to sample images in the polar coordinate system and map the samples to Cartesian space for construction of new (augmented) images. A radial transformed image is a coherent representation of the original image and maintains the semantic validity of the data classes.

2. PROPOSED METHOD

We define a point on a plane in the polar coordinate system as $(r, \theta) \in \mathcal{P}^2$ where $r \in \mathbb{Z}^+$ is the radial coordinate and

(a) Sampling pixels using radial transform.

(b) Constructed image from selected samples in the polar coordinate system.

(c) Ring enhancing mass on a brain MRI.

(d) Augmented MRI from original MRI using radial transform.

Fig. 1: Radial transform sampling. a) Selected discrete samples using radial transform on an arbitrary plane. The arbitrary selected pole is at pixel $O(170, 50)$; b) Mapping of selected samples in (a) from polar coordinate system to Cartesian coordinate system. The red samples show the direction of mapping the samples in (a) into (b); c) An example of radial transform application in tumor detection: Ring enhancing mass on an original brain MRI. Detection of the mass requires non-linear threshold(s); d) The original brain MRI in (c) after radial transform with pole selected in the middle of the mass. The ring enhancing mass is expanded (up-sampled) and can be detected with a linear threshold.

$\theta \in \mathbb{R}^+$ is the counterclockwise angular coordinate with respect to a polar axis drawn horizontally from the pole to the right, as illustrated in Figure 1(a). We draw M distinct rays with length $R_m \in \mathbb{R}^+$ where $m = 0$ is the arbitrary initial ray along the polar axis. Each ray has an identical angular difference $\Delta\theta = |\theta_{m+1} - \theta_m|$ from its adjacent ray. Then, we generate a set of spatial coordinates $K = \{(r, \theta_m)_k\}$ in $\mathcal{P}^2$, with respect to a pole $O(u, v) \in \mathcal{P}^2$ such that $u, v \in \mathbb{Z}^+$, which is an arbitrary selected pixel in the original image $\mathbf{X}_{M \times N} \in \mathcal{C}^2$ in the Cartesian coordinate system. Therefore, to generate a point $(r, \theta_m)_k$ for $r \in \{0, ..., N - 1\}$ we have

$$\theta_m = 2\,\pi \cdot m/M, \tag{1}$$

where $m \in \{0, ..., M - 1\}$. By $\mathbf{X} \xrightarrow{\phi(r, \theta_m)_k} \hat{\mathbf{X}}$, we map the pixels at Cartesian coordinates $(m, n) \in \mathbb{Z}^+$ from the original image $\mathbf{X}$ in $\mathcal{C}^2$ to construct the augmented image $\hat{\mathbf{X}} \in \mathcal{C}^2$ in the Cartesian space using radial transform $\phi(\cdot)$ as

$$\hat{x} = round(r \cdot cos(\theta_m)) \ \& \ \hat{y} = round(r \cdot sin(\theta_m)) \tag{2}$$

for $r \in \{0, ..., N - 1\}$ and $m \in \{0, ..., M - 1\}$ such that $0 \le u + \hat{x} < M$ and $0 \le v + \hat{y} < N$. These conditions guarantee that the pair $(\hat{x}, \hat{y})$ stays spatially within $\hat{\mathbf{X}}$. A new pixel (m, r) in the constructed image is then defined as $\hat{x}_{m,r} = x_{u+\hat{x}, v+\hat{y}}$. The image $\tilde{\mathbf{X}}$ is the radial transform of $\mathbf{X}$ with respect to the pixel $O(u, v) \in \mathbf{X}$.

The pole at pixel location $O(u, v)$ in $\mathbf{X}$ is repeated as the first pixel in every row $m \in \{0, ..., M - 1\}$ of $\hat{\mathbf{X}}$, as illustrated in Figure 1(b). As $r \to N$ for an arbitrary θ_m, pixels in close neighborhood of $O(u, v)$ are up-sampled and pixels in further neighborhood are down-sampled. The proposed radial transfrom can generate unique radial transformed images up to the number of pixels in $\mathbf{X}$, which is $M \times N$. Such high diversity of images preserves the dependencies among local and global pixels in $\mathbf{X}$ but in different representations. For a typical 256×256 original image, the rotation augmentation approach generates much less new representations of the original image (e.g., 360 new representations by having a rotation step of one degree) comparing with the radial transform, which generates $256 \times 256 = 65,536$ new images.

Figure 1(c) shows a practical example on how radial transform can help to map a set of pixels belonging to a ring enhancing mass in a brain magnetic resonance imaging (MRI) to be mapped into a new image using polar coordinate system. The pole is selected in the middle of the mass. Distinguishing the mass in the Cartesian plane requires a set of non-linear thresholds. However, radial transform in Figure 1(d) has expanded (i.e., up-sampled) the image in local neighborhood of the mass and has compressed (down-sampled) the further, less important pixels. In this way the mass can be easily separated with a linear threshold. This example clearly shows how radial transform not only can help to augment images but also can reduce complexity of classification problems in images. In this paper, our focus is on the augmentation property of the radial transform.

A neural network can be trained with the generated radial transformed images. An image can be classified by applying $\phi(\cdot)$ on a subset of pixels T and detect the corresponding classification labels $L = \{c_t : c_t = argmax\{p_{t,1}, ..., p_{t,C}\} \ \forall \ t \in T\}$ using a trained network with radial transformed images. These labels can be directly used for segmentation and multi-object detection applications. As an example for the single-object image classification task, the majority of votes from the predicted labels (i.e., the most frequent label in L) is the predicted class

3. EXPERIMENTS

Experiments were conducted using GoogLeNet [10] and AlexNet [5] on two datasets: the MNIST dataset with 10 classes of hand written digits [11] and a dataset of naturally limited 9 different modalities of medical images [12]. A sample from some of the image classes is presented in the first row of Figure 2.
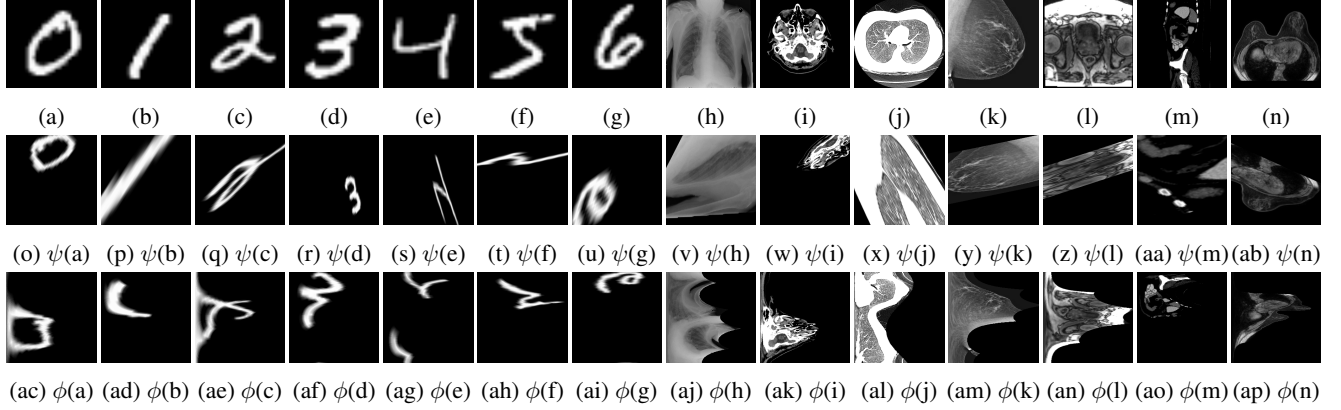
Fig. 2: A sample from some data classes of MNIST and medical multimodal datasets with their corresponding representation using radial transform, $(\phi(\cdot))$ and Affine transform $(\psi(\cdot))$, for a randomly selected pixel. a) 0; b) 1; c) 2; d) 3; e) 4; f) 5; g) 6; h) Chest X-Ray; i) Head CT; j) Lung CT; k) Mammogram; l) Pelvis MRI; m) Sagittal Abdomen MRI; n) Breast MRI.

3.1. Settings

In order to evaluate how the radial transform can increase diversity of a dataset and help the deep learning model to improve the generalization performance, the standard AlexNet and GoogLeNet models are trained with three different datasets: D1) Original images with 20 and 40 samples per class; D2) Affine transformed $\psi(\cdot)$ images of the original images, where 100 images are made from each original image, totally two datasets of size 2,000 and 4,000 images per class. The transformation parameters are selected randomly; D3) Radial transformed $\phi(\cdot)$ images of the original images, where 100 images are made from each original image, totaling two datasets of size 2,000 and 4,000 images per class. The pole is selected randomly in the original image. The datasets are made for MNIST and medical multimodal images, totaling six datasets.

The AlexNet and GoogLeNet models are trained using Stochastic Gradient descent over 50 iterations with exponential decay learning rate initialized to 0.01 and 0.001 for models trained with original and radial or Affine transformed images, respectively. The parameters are selected based on grid search. The training dataset is shuffled to avoid sensitivity of the models to training order. The models are cross-validated over 30 independent experiments and a statistical test is conducted on the results. The validation and test datasets have 1,000 images each, with identical number of samples per class.

3.2. Results Analysis

Figure 2 shows how the Affine transform and the proposed radial transform can augment an image to generate a new representation. The augmented images using Affine transform show this transform can preserves points, maps a line to a line, and preserves parallel lines such as the head and tail of the number 5's image, ribs in the chest X-ray, and left and right obturator internus of pelvis MRI. This transform

also preserves ratios of distances between points lying on a straight line. For example, the distance between the two sides (branches) of the number 4's image at top and bottom, fatty tissues in the mammogram, and Cerebellar hemisphere in the head computed tomography (CT). However, this transform may result in loss of resolution or a part of image as shown for the head CT, and chest X-ray. The resulted images from radial transform show that it preserves the local and global spatial features in the neighborhood of the pole. This transform up-samples the pixels sitting in close spatial proximity of the pole and down-samples the pixels distant to the spatial neighborhood of the pole. Despite the Affine transform, the radial transform does not necessarily preserve the parallel lines or distances between two specific point. However, it defines a logical relationship among the pixels based on the sampling in polar coordinate system.

The accuracy (υ), top-one probability confidence (κ), and converged-in iteration (ξ) of AlexNet and GoogLeNet are presented in Table 1. The performance values per image class for original dataset size of 20 and corresponding Affine and radial transformed images are presented in Tables 2 and 3 (due to lack of space, the datasize of 20 is only presented). The accuracy of class c is defined as $\upsilon_c = \sum_{s=1}^{|S|} \mathbb{I}[c = arg\ max(p_{s,1}, ..., p_{s,C})]/|S|$ where S is the test dataset, C is the number of classes, $p_{c,s}$ is the classification probability of the data sample s for class c, and $\mathbb{I}[x]$ is defined to be 1 if x is true, and 0 if it is false. The top-one probability confidence of class c is defined as $\kappa_c = (\sum_{s=1}^{|S|} p_{s,c})/|S|$.

The results clearly show that the models trained with radial transformed data have better performance. At competitive accuracy, the confidence of models trained with radial transform is greater. The difference in accuracy of the trained models is more obvious for the MNIST dataset, likely due to the correlation among the medical images such as between Transverse abdomen MRI and Sagittal abdomen MRI.

The accuracy of the model on the validation dataset

Table 1: The accuracy ("υ" in %), top-one probability confidence value ("κ" in %), and converged-in iteration (ξ) of AlexNet and GoogLeNet models trained with original, Affine transform, and radial transform augmented MNIST and medical multimodal images. "Std" is the standard deviation. The best result is in boldface.

Model	Transform	Number of Original Images per Class											
		MNIST						Medical Multimodal					
		20			40			20			40		
		υ	κ	ξ	υ	κ	ξ	υ	κ	ξ	υ	κ	ξ
GoogLeNet	Original	45.62±1.91	28.67±2.28	48	69.39±0.73	67.00±0.69	42	83.00±2.51	78.93±2.32	40	98.33±0.81	95.95±0.81	26
	Affine	11.16±0.10	12.86±0.01	2	59.11±0.56	56.88±0.61	19	48.28±0.72	55.21±0.70	12	61.42±1.02	58.92±0.95	16
	Radial	97.98±1.39	99.39±1.22	13	91.96±0.89	92.72±0.93	15	94.48±1.36	97.33±1.42	8	99.21±0.92	99.14±0.98	8
AlexNet	Original	82.26±1.06	84.35±1.28	33	83.60±0.51	83.30±0.56	40	89.01±1.28	88.14±1.39	21	98.04±0.61	98.33±0.58	25
	Affine	32.52±0.43	33.54±0.48	43	38.50±0.52	37.17±0.59	48	52.80±0.78	46.61±0.89	15	47.66±0.82	43.46±0.89	7
	Radial	**98.29±0.96**	**98.57±0.98**	19	**95.18±0.74**	**94.36±0.69**	17	**97.05±1.01**	**99.34±1.21**	4	**99.54±0.66**	**98.01±0.51**	4

Table 2: The accuracy ("υ" in %) and top-one probability confidence value ("κ" in %) of AlexNet and GoogLeNet models trained with original, Affine transform, and radial transform augmented medical multimodal images. The dataset size is 20 and 2,000 per class for the original and transformed images, respectively.

Model	Transform	Category																	
		Coronal Abd.		Trans. Abd.		Sagittal Abd.		Breast MRI		Chest X-Ray		Head CT		Lung CT		Mammogram		Pelvis MRI	
		υ	κ	υ	κ	υ	κ	υ	κ	υ	κ	υ	κ	υ	κ	υ	κ	υ	κ
GoogLeNet	Original	92.18	78.21	89.99	89.98	30.78	53.12	94.44	67.13	75.24	89.13	75.24	89.13	96.24	72.78	98.03	88.19	94.87	82.70
	Affine	1.00	13.73	1.00	16.83	48.00	52.26	95.82	83.94	79.22	70.29	82.53	78.49	27.00	52.34	65.00	74.83	35.00	48.83
	Radial	98.12	99.59	**99.90**	94.92	82.53	86.18	97.25	99.15	85.17	97.84	**90.17**	99.37	99.33	99.14	**99.79**	99.82	98.06	99.99
AlexNet	Original	96.00	91.51	95.10	97.00	54.28	89.25	97.00	99.97	93.18	92.69	87.91	89.07	92.37	91.80	94.03	90.04	91.30	90.87
	Affine	1.00	12.67	0.00	0.00	69.00	65.22	96.03	84.20	71.45	69.19	81.77	85.28	54.00	66.29	68.00	73.92	34.00	52.91
	Radial	**98.96**	**99.60**	99.70	**99.91**	**86.31**	**95.38**	**99.90**	**99.98**	**99.99**	**99.99**	89.17	**99.38**	**100.00**	**99.99**	99.49	**99.91**	**100.00**	**99.99**

Table 3: The accuracy ("υ" in %) and top-one probability confidence value ("κ" in %) of AlexNet and GoogLeNet models trained with original, Affine transform, and radial transform augmented MNIST images. The dataset size is 20 and 2,000 per class for the original and transformed images, respectively.

Model	Transform	Category																		
		0		1		2		3		4		5		6		7		8		9	
		υ	κ	υ	κ	υ	κ	υ	κ	υ	κ	υ	κ	υ	κ	υ	κ	υ	κ	υ	κ
GoogLeNet	Original	21.29	22.60	96.14	89.98	16.69	19.57	48.44	18.84	66.22	33.27	23.49	17.22	48.08	17.59	72.93	30.17	25.96	16.96	37.04	20.57
	Affine	0.54	13.19	90.00	13.86	1.00	17.30	2.47	16.83	1.25	17.92	3.12	15.24	3.01	15.93	0.00	0.00	1.20	17.91	1.00	19.30
	Radial	98.25	**99.35**	97.24	98.98	98.22	**99.41**	99.09	**99.74**	98.40	**99.61**	96.92	**99.37**	98.98	99.59	97.60	**99.25**	**97.64**	99.46	97.54	99.23
AlexNet	Original	83.65	90.01	96.24	95.77	79.62	85.57	60.19	71.31	92.67	89.71	59.99	74.45	89.55	84.67	90.73	91.74	82.65	77.36	87.30	82.87
	Affine	57.75	46.81	87.46	40.15	9.78	46.00	21.79	57.47	26.55	30.03	12.32	47.24	17.16	35.06	26.65	41.25	33.11	37.62	32.70	45.24
	Radial	**99.37**	99.02	**98.76**	**99.21**	**98.45**	98.38	**99.52**	99.26	**99.22**	98.74	**97.34**	98.37	98.29	98.98	**97.90**	98.00	96.62	97.93	97.42	97.81
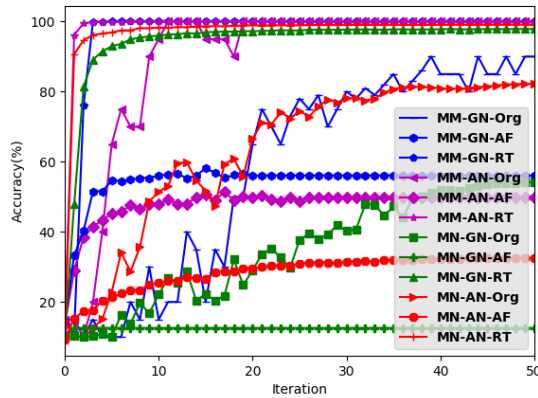
Fig. 3: Accuracy of the AlexNet (AN) and GoogLeNet (GN), trained with original (Org), Affine transformed (AF), and radial transformed (RT) images on MNIST (MN) and medical multimodal (MM) validation datasets.

through training iterations for a single experiment is presented in Figure 3. The converged-in iteration ξ in Table 1 and this figure show that the models trained with radial transformed images D3 converge faster with higher accuracy than models trained with D1 and D2. AlexNet and GoogLeNet trained with very limited original images show fluctuation of validation accuracy during training. GoogLeNet has more fluctuation, particularly due to having more number of free parameters than AlexNet. The same models trained with augmented images using radial transform show smoother convergence and less fluctuation of the validation accuracy.

4. CONCLUSION

Successful training of deep neural networks requires a large quantity of balanced data. In practice, most of the datasets are imbalanced and often very limited data is available for certain classes in a dataset. In this paper, we propose image augmentation using radial transform in the polar coordinate system to facilitate training of deep neural networks. This method preserves the information content of the original image, but improves the diversity of the training dataset, resulting in improved generalization performance of the neural network.

5. REFERENCES

[1] Zhi-Hua Zhou and Xu-Ying Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, 2006.

[2] Haibo He and Edwardo A Garcia, "Learning from imbalanced data," *IEEE Transactions on knowledge and data engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[3] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[4] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting.," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[6] Justin Salamon and Juan Pablo Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, 2017.

[7] Charles C Stearns and Karthikeyan Kannappan, "Method for 2-d affine transformation of images," Dec. 12 1995, US Patent 5,475,803.

[8] Emmanuel Okafor, Rik Smit, Lambert Schomaker, and Marco Wiering, "Operational data augmentation in classifying single aerial images of animals," in *INnovations in Intelligent SysTems and Applications (INISTA), 2017 IEEE International Conference on*. IEEE, 2017.

[9] Pew-Thian Yap, Xudong Jiang, and Alex Chichung Kot, "Two-dimensional polar harmonic transforms for invariant image representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1259–1270, 2010.

[10] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[11] Yann LeCun, Corinna Cortes, and Christopher JC Burges, "Mnist handwritten digit database," *AT&T Labs [Online]. Available: http://yann. lecun. com/exdb/mnist*, vol. 2, 2010.

[12] "The image data in the cancer imaging archive (tcia)," *http://www.cancerimagingarchive.net/*.