Training Neural Networks with Very Little Data- A Draft

Hojjat Salehinejad, Joseph Barfett, Shahrokh Valaee, *Senior Member, IEEE*, and Timothy Dowdell

***Abstract*—Deep neural networks are complex architectures composed of many layers of nodes, resulting in a large number of parameters including weights and biases that must be estimated through training the network. Larger and more complex networks typically require more training data for adequate convergence than their more simple counterparts. The data available to train these networks is often limited or imbalanced. We propose the radial transform in polar coordinate space for image augmentation to facilitate the training of neural networks from limited source data. Pixel-wise coordinate transforms provide representations of the original image in the polar coordinate system and both augment data as well as increase the diversity of poorly represented classes. Experiments performed on MNIST and a set of multimodal medical images using the AlexNet and GoogLeNet neural network models show high classification accuracy using the proposed method.**

***Index Terms*—Imbalanced data, medical images, neural networks, polar coordinate system, radial transform, sampling.**

I. Introduction

DEEP learning refers to a network of connected artificial neurons in multiple layers, which can perform feature extraction from observed data and learn the complicated relationships among the features of data. These models generally use non-linear but simple units, which can provide different levels of feature representation through their multi-layer architecture. The higher layers provide a more abstract representation of data and suppress irrelevant variations [1]. The extracted features are later passed to a classifier. Deep networks learn by observing various examples as input data and minimizing the error in classifying the corresponding labels.

The need for massive amounts of data to train deep neural networks is a major drawback to these models. Accuracy, performance, and versatility of deep learning models are highly dependent on availability of abundant data. The other challenge is imbalanced datasets, where very few data samples are available for some data classes. Data collection itself is a complex and expensive process for large-scale experiments and can include challenges in ethics board approval and patient privacy. These challenges arise in many practical machine learning scenarios such as fraud detection in banking/marketing transactions [2] or in medical sciences [3]. 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 network.

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 this paper, the mapping of 2-dimensional image data from the Cartesian coordinate system to the polar coordinate system using a radial transform as a means of exponential data augmentation is proposed. A radial transformed image is a coherent representation of the original image and it is possible to apply such a transformation to produce additional training data while maintaining the semantic validity of the labels.

II. Background

Neural networks trained with little data are potentially sensitive to training order and parameter initialization. In training, the optimization algorithm usually leverages data randomness to converge upon optimal weights and biases of the network. The combination of such properties can result in an "unstable" or "under-fit" network prone to sporadic outputs, especially in the early iterations of training. In some cases, it is possible to achieve reasonable performance by cross-validation of the model's hyper-parameters [7].

In general, data augmentation does not increase the information content of the dataset. However, it can improve diversity of the dataset and generalization performance. The generalization performance refers to the accuracy of the neural network in classification of unseen data. 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 transformation matrix, does not affect the semantic meaning of the labels [6].

In audio augmentation, the combination of deformation parameters such as time stretching and pitch shifting must maintain the semantic validity of the label [6]. An audio signal

H. Salehinejad is with the Department of Electrical & Computer Engineering, University of Toronto, Toronto, Canada, and Department of Medical Imaging, St. Michael's Hospital, University of Toronto, Toronto, Canada, e-mail: salehinejadh@smh.ca.

J. Barfett is with the Department of Medical Imaging, St. Michael's Hospital, University of Toronto, Toronto, Canada, e-mail: barfettj@smh.ca.

S. Valaee is with the Department of Electrical & Computer Engineering, University of Toronto, Toronto, Canada, e-mail: valaee@ece.utoronto.ca.

T. Dowdell is with the Department of Medical Imaging, St. Michael's Hospital, University of Toronto, Toronto, Canada, e-mail: dowdellt@smh.ca.

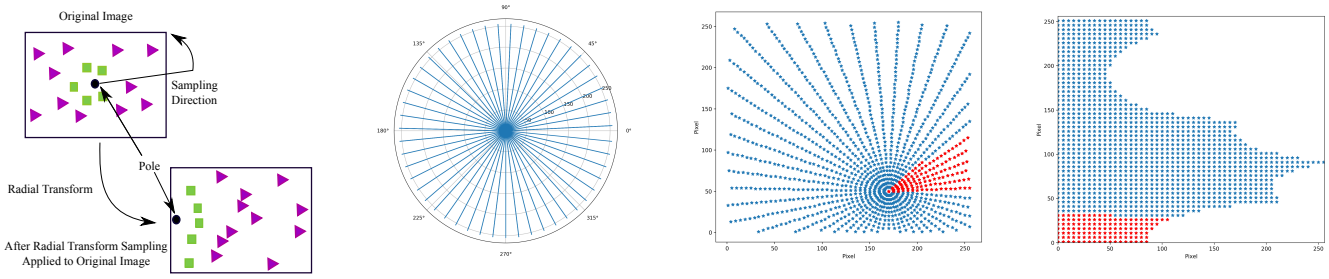


Fig. 1: Radial transform sampling. a) Mapping samples from Cartesian coordinate system (left) into a polar coordinate system (right) using the radial transform. b) Radial transform in polar coordinate system. c) Selected discrete samples of a 256×256 image (2D plane) using radial transform. The arbitrary selected pole is at pixel (170,50). d) Mapping of selected samples in (c) from polar coordinate system to Cartesian coordinate system. The red samples show the direction of mapping the samples in (c) into (d).

Algorithm 1 Radial Transform (RT)

Procedure RT($\mathbf{I}, U, V, m, n$)
$\tilde{\mathbf{I}}_{U \times V} = 0$ //matrix of zeros
for $u = 0 \rightarrow U - 1$ **do**
  $\theta_u = 2\,\pi \cdot u/U$
  for $v = 0 \rightarrow V - 1$ **do**
    $x = \|v \cdot cos(\theta_u)\|$
    $y = \|v \cdot sin(\theta_u)\|$
    if $0 \leq m + x < U$ & $0 \leq n + y < V$ **then**
      $\tilde{\mathbf{I}}_{u,v} = \mathbf{I}_{m+x,n+y}$
    end if
  end for
end for
return $\tilde{\mathbf{I}}$

can be processed as an image by plotting the frequencies of the signal over a time axis and brightness or color representing the strength of a frequency component at each time frame, referred to as spectrogram [8]. Music auto-tagging is one example where the tags are highly diverse and have different levels of abstraction [8]. The spectrogram of music can be fed as an image to embrace multi-level and multi-scaled features [8].

Some of the image data augmentation techniques include adding noise, rotating, translating, mirroring, or scaling the image. Another approach is making an image that contains multiple copies of the original image rotated by different angles [9]. Spherical images augmented with depth information and a spherical saliency map were developed for vision-based navigation in real urban environments [10]. The polar harmonic transform, based on a set of orthogonal projections, has been used to generate a set of features that are insensitive to rotation [11]. The log-polar transform has been used for image registration [12], [13], whereby the transform can eliminate the rotation and scale effects on an image by constructing a new representation in the polar coordinate system. Hierarchical kernels then can eliminate the row and column translation effects from the new image [14].

One of the most efficient classification methods in computer vision is convolutional neural network (CNN). CNNs are used for various applications including human skin detection [15], environmental sound classification [6], light field image super-resolution [16], and single image super-resolution [17]. A typical CNN uses a pooling strategy to represent the outputs of surrounding groups of neurons feeding a common kernel map. AlexNet is one of the advanced CNN models where the neighborhoods summarized by adjacent pooling units overlap to help the model become resistant to over-fitting [5].

The architectures of GoogLeNet and AlexNet share similar fundamental features [5], [18]. GoogLeNet achieves an improved utilization of computing resources by implementing the Hebbian principle and intuition of multi-scale processing. This approach increases the potential depth and width of a the network with fixed computational complexity [18].

III. The Proposed Method

The polar coordinate system describes a point on a 2-dimensional plane as the distance of the point from the reference point (i.e., the pole) and an angle from an arbitrary reference angle. Figure 1a shows how a set of points belonging to two different categories in the Cartesian plane can be mapped into a polar coordinate system. Distinguishing the samples belonging to these two data classes in the Cartesian plane requires a non-linear classifier and takes more effort than the required linear classifier in a polar coordinate system for the same set of data.

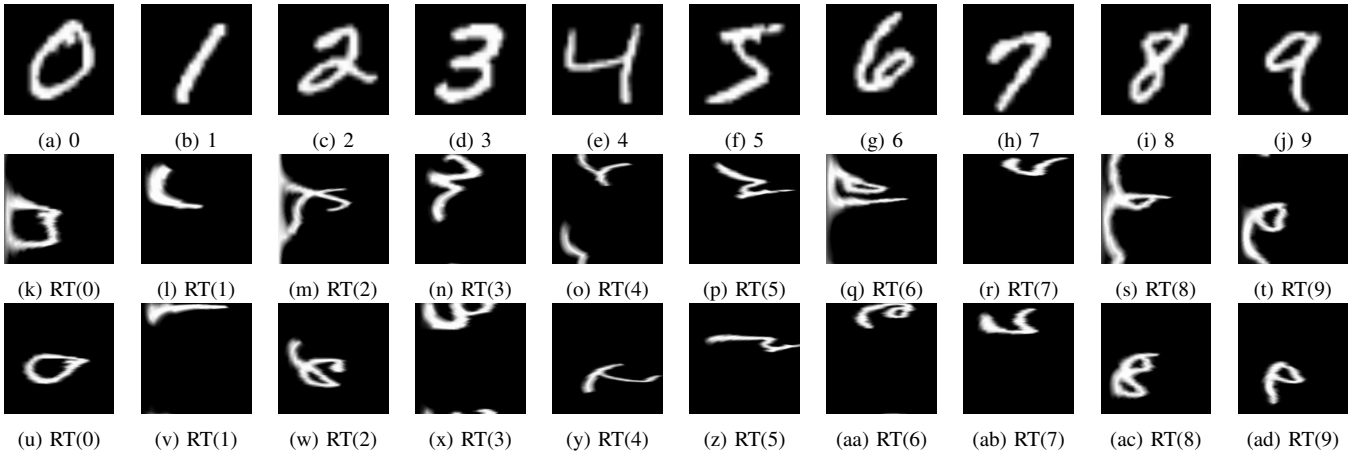
In this paper, we develop a similar mapping concept from the Cartesian to the polar coordinate system for sampling an image. In the polar coordinate system, we can draw U distinct rays with length V where $u = 0$ is the arbitrary initial ray. Each ray has an identical angular difference θ from its adjacent ray, as illustrated in Figure 1b. The angular difference of ray u from the initial reference ray $u = 0$ is defined as

$$\theta_u = 2\,\pi \cdot u/U \quad \forall\, u \in \{0, ..., U-1\}. \tag{1}$$

Therefore, an arbitrary point on the ray u with distance v from the pole and angular difference θ_u can be represented with the pair (θ_u, v).

On an image $\mathbf{I}$ with size $U \times V$ we can select an arbitrary pixel at Cartesian coordinate (m, n) as the pole in the polar coordinate system. For example, we can select the pixel $(m, n) = (170, 50)$ in the Figure 1c, where $U = V = 256$. For $v \in \{0, ..., V-1\}$ and $u \in \{0, ..., U-1\}$, each ray can visit the pixels at

$$x = \|v \cdot cos(\theta_u)\| \text{ and } y = \|v \cdot sin(\theta_u)\| \tag{2}$$

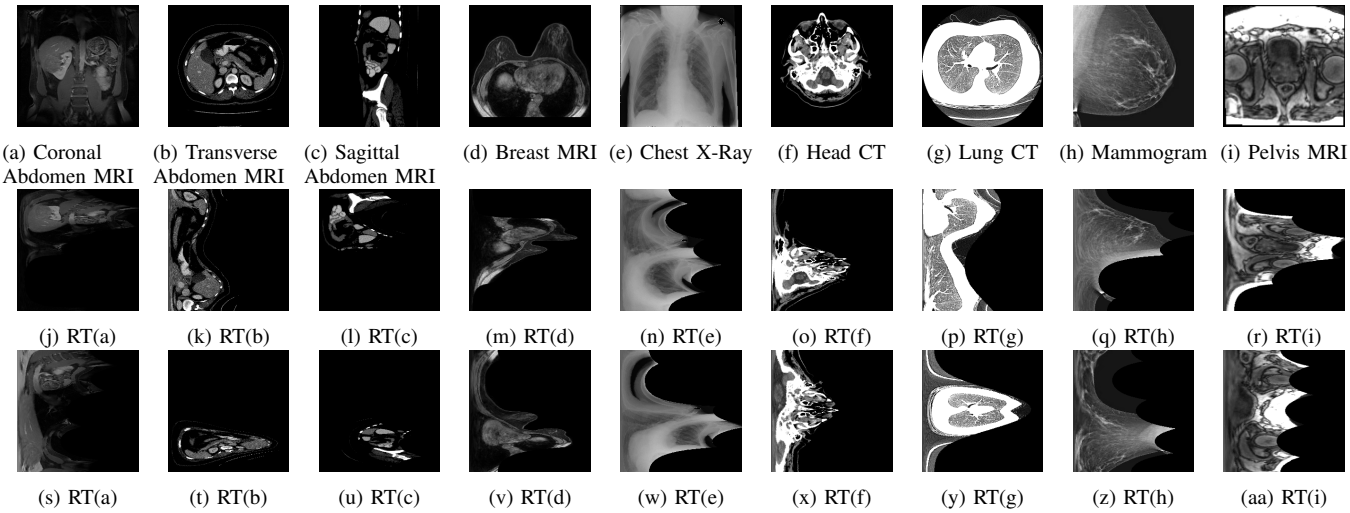


(a) 0 (b) 1 (c) 2 (d) 3 (e) 4 (f) 5 (g) 6 (h) 7 (i) 8 (j) 9

(k) RT(0) (l) RT(1) (m) RT(2) (n) RT(3) (o) RT(4) (p) RT(5) (q) RT(6) (r) RT(7) (s) RT(8) (t) RT(9)

(u) RT(0) (v) RT(1) (w) RT(2) (x) RT(3) (y) RT(4) (z) RT(5) (aa) RT(6) (ab) RT(7) (ac) RT(8) (ad) RT(9)

Fig. 2: Samples from the MNIST dataset and corresponding representations using radial transform RT(·) in polar coordinate system.



(a) Coronal Abdomen MRI (b) Transverse Abdomen MRI (c) Sagittal Abdomen MRI (d) Breast MRI (e) Chest X-Ray (f) Head CT (g) Lung CT (h) Mammogram (i) Pelvis MRI

(j) RT(a) (k) RT(b) (l) RT(c) (m) RT(d) (n) RT(e) (o) RT(f) (p) RT(g) (q) RT(h) (r) RT(i)

(s) RT(a) (t) RT(b) (u) RT(c) (v) RT(d) (w) RT(e) (x) RT(f) (y) RT(g) (z) RT(h) (aa) RT(i)

Fig. 3: Samples from the multimodal medical dataset and corresponding representations using radial transform RT(·) in polar coordinate system.

such that $0 \leq m + x < U$ and $0 \leq n + y < V$. These conditions guarantee that the pair (x, y) stays in within the image. A new image $\tilde{\mathbf{I}}$ using the mapping illustrated in Figure 1a can be constructed by taking the visited pixel (x, y) and positioning it such that $\tilde{\mathbf{I}}_{u,v} = \mathbf{I}_{m+x,n+y}$. The image $\tilde{\mathbf{I}}$ is the radial transform of $\mathbf{I}$ in the polar coordinate system. The radial transform procedure is summarized in Algorithm 1.

The pole at pixel location (m, n) in the original image $\mathbf{I}$ is repeated as the first pixel in every row $u \in \{0, ..., U-1\}$ of the radial transformed image $\tilde{\mathbf{I}}$. As $v \to V$ for an arbitrary θ, pixels close to the pole (m, n) are over-sampled and pixels away from the pole are under-sampled or missed. This technique constructs $U \times V$ different radial transformed images of size $U \times V$ from a single image $\mathbf{I}_{U \times V}$. Such high diversity of images preserves the dependencies among local and global pixels in the original image $\mathbf{I}$ but in different representations.

A neural network can be trained with the generated radial transformed images. An original test image $\bar{\mathbf{I}}$ can be classified by applying radial transform on a subset of its pixels T and receiving the corresponding classification labels $L = \{c_t : c_t = arg\ max\{p_{t,1}, ..., p_{t,C}\} \forall t \in T\}$ from the trained network. These labels can be directly used for segmentation and multi-object detection applications. For single-object image classification, the majority of votes from the predicted labels (i.e., the most frequent label in L) is the predicted class of the image.

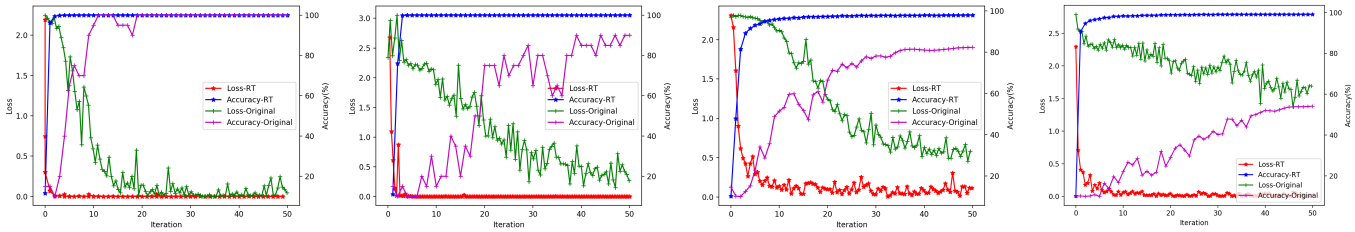
IV. Experiments

Experiments are conducted using AlexNet [5] and GoogLeNet [18] on the MNIST dataset with 10 classes of hand written digits [19] and a dataset of 9 different medical images modalities [20], as presented in Figures 2 and 3, respectively.

A. Settings

AlexNet and GoogLeNet are trained once with a small dataset (i.e., 20 images per class) and once with 100 radial transformed images generated from each image (i.e., 2,000 radial images per class) of the same dataset. The number of images per class is selected such that it represents a very small dataset. For example, since the MNIST dataset has 10 classes, the original truncated dataset has 200 images and the radial transformed dataset has 20,000 images.

The models are trained using Stochastic Gradient descent with mini-batch size of 64 over 50 iterations with exponential decay learning rate initialized to 0.01 and 0.001 for models trained with original and radial transformed images, respectively. The parameters are selected based on grid search. The



(a) AlexNet-Multimodal (b) GoogLeNet-Multimodal (c) AlexNet-MNIST (d) GoogLeNet-MNIST

Fig. 4: Convergence behavior of the AlexNet and GoogLeNet models trained with the original and radial transformed images in MNIST and medical multimodal datasets. The term "RT" refers to radial transformed images and the term "Original" refers to models trained with very little original images. The x-axis represents training iteration, left y-axis represents loss of model while training, right y-axis represents the accuracy of the model while training, using the validation dataset.

TABLE I: The accuracy ("Acc." in %) and confidence value ("Conf." in %) of AlexNet and GoogLeNet trained with original and radial transformed multimodal medical images. "Abd" refers to abdominal MRI and "Std" is the standard deviation. The best result is in boldface.

Model	Data	Category																		Total	
		Coronal Abd.		Transverse Abd.		Sagittal Abd.		Breast MRI		Chest X-Ray		Head CT		Lung CT		Mammogram		Pelvis MRI			
		Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.±Std	Conf.±Std
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	92.46±2.78	88.14±2.89
	Radial	98.96	99.60	99.70	99.91	86.31	95.38	99.90	99.98	99.99	99.99	89.17	99.37	100.00	99.99	99.49	99.91	100.00	99.99	**97.05±1.01**	**99.34±1.21**
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	83.00±2.91	78.93±3.02
	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	**94.48±1.36**	**97.33±1.42**

TABLE II: The accuracy ("Acc." in %) and confidence value ("Conf." in %) of AlexNet and GoogLeNet trained with original and radial transformed MNIST images. "Std" is the standard deviation. The best result is in boldface.

Model	Data	Category																		Total	
		0		1		2		3		4		5		6		7		8		9			
		Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.	Conf.	Acc.	Conf.
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	82.26±2.06	84.35±2.28
	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	**98.29±0.96**	**98.57±0.98**
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	45.62±3.11	28.67±3.28
	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	**97.98±1.39**	**99.39±1.22**

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 each has 1,000 images.

B. Results Analysis

The accuracy and confidence values of AlexNet and GoogLeNet trained with original and radial transformed test images are presented in Tables I and II. The accuracy for class c in percentile is defined as $100 \times (\sum_{s=1}^{|S|}[c = arg\ max(p_{s,1}, ..., p_{s,C})])/|S|$ 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 $[x]$ is defined to be 1 if x is true, and 0 if it is false. The confidence for class c is defined in percentile as $100 \times (\sum_{s=1}^{S} p_{s,c})/S$. The results clearly show that the models trained with radial transformed data have higher accuracy and greater confidence value. At competitive accuracy, the confidence of models trained with radial transformation is greater. The difference in accuracy of the trained models with original data and radially transformed data is more obvious for the MNIST dataset, likely due to the correlation among the medical images such as between Transverse abdomen magnetic resonance imaging (MRI) and Sagittal abdomen MRI.

The performance of AlexNet is better compared with the GoogLeNet for training with a small dataset. The convergence behavior of the trained models with the original and radial transformed images (indicated with RT) for a single experiment is presented in Figure 4. The RT models perform training for multiple mini-batches of images per iteration and converge faster than models trained with small quantities of conventional data. The models trained with images generated using the radial transformation have higher accuracy and less loss.

The models trained with very limited original images show fluctuation of training loss and validation accuracy in Figure 4. This is while the same models, but trained with augmented images using radial transformation, show smooth convergence of the training loss and validation accuracy values.

V. 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 or for a study. In this paper, we propose the radial transform in the polar coordinate system to increase the number of samples in the dataset to facilitate the training of deep neural networks. The proposed data augmentation method does not change the information content of the data but improves the diversity of the data. Our results show that this

approach increases generalization performance of the neural networks, defined as the accuracy of machine learning model in predicting outcome values for previously unseen data. Training the state-of-the-art AlexNet and GoogLeNet deep neural network models on very little data shows fluctuation of training loss and validation accuracy throughout the learning process.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] H. Salehinejad and S. Rahnamayan, "Customer shopping pattern prediction: A recurrent neural network approach," in *Computational Intelligence (SSCI), 2016 IEEE Symposium Series on*. IEEE, 2016, pp. 1–6.

[3] F. Pouladi, H. Salehinejad, and A. M. Gilani, "Recurrent neural networks for sequential phenotype prediction in genomics," in *Developments of E-Systems Engineering (DeSE), 2015 International Conference on*. IEEE, 2015, pp. 225–230.

[4] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. 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] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[6] J. Salamon and J. P. 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] D. Maclaurin, D. Duvenaud, and R. Adams, "Gradient-based hyperparameter optimization through reversible learning," in *International Conference on Machine Learning*, 2015, pp. 2113–2122.

[8] J. Lee and J. Nam, "Multi-level and multi-scale feature aggregation using pre-trained convolutional neural networks for music auto-tagging," *IEEE Signal Processing Letters*, vol. 24, pp. 1208–1212, 2017.

[9] E. Okafor, R. Smit, L. Schomaker, and M. Wiering, "Operational data augmentation in classifying single aerial images of animals."

[10] M. Meilland, A. I. Comport, and P. Rives, "A spherical robot-centered representation for urban navigation," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 5196–5201.

[11] P.-T. Yap, X. Jiang, and A. C. 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.

[12] R. Matungka, Y. F. Zheng, and R. L. Ewing, "Image registration using adaptive polar transform," *IEEE Transactions on Image Processing*, vol. 18, no. 10, pp. 2340–2354, 2009.

[13] J. Zhang, F. Sohel, H. Bian, M. Bennamoun, and S. An, "Forward-looking sonar image registration using polar transform," in *OCEANS 2016 MTS/IEEE Monterey*. IEEE, 2016, pp. 1–6.

[14] Y. Y. Tang, T. Xia, Y. Wei, H. Li, and L. Li, "Hierarchical kernel-based rotation and scale invariant similarity," *Pattern Recognition*, vol. 47, no. 4, pp. 1674–1688, 2014.

[15] H. Zuo, H. Fan, E. Blasch, and H. Ling, "Combining convolutional and recurrent neural networks for human skin detection," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 289–293, 2017.

[16] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. S. Kweon, "Light-field image super-resolution using convolutional neural network," *IEEE Signal Processing Letters*, vol. 24, no. 6, pp. 848–852, 2017.

[17] J. Y. Cheong and I. K. Park, "Deep cnn-based super-resolution using external and internal examples," *IEEE Signal Processing Letters*, 2017.

[18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[19] Y. LeCun, C. Cortes, and C. J. Burges, "Mnist handwritten digit database," *AT&T Labs [Online]. Available: http://yann. lecun. com/exdb/mnist*, vol. 2, 2010.

[20] [Online]. Available: http://www.cancerimagingarchive.net/