

Classification of Infiltrated Injections During PET/CT Imaging Applying Deep Learning Technique

Tasmia Rahman Tumpa¹, Shelley N. Acuff¹, Dustin R. Osborne¹

Abstract—Objective: Injected dose infiltration can negatively impact quantitative evaluation of Positron Emission Tomography (PET) data by leading to inaccurate calculation of Standardized Uptake Value (SUV) measurements and limiting bioavailability of the tracer in the patient. Recently developed topical gamma scintillation sensors provide a way to monitor Time Activity Curves (TACs) and determine the presence of activity remaining at the injection site after injection. However, TAC analysis and visual inspection by physician of static PET images differ in many cases which has been a recent research concern. In this work, a deep learning (DL) based classification was implemented to study whether this approach can be a viable solution to classify good/infiltrated data. **Method:** A supervised machine learning technique was adopted and TACs obtained from the sensors were fed as input to a neural network. The network was trained to classify two classes of data i.e. good quality injections and poor quality injections. The performance of the network was tested on the basis of 3-fold cross-validation. **Result:** The network could label the good injection data (94.58% data) with around $\approx 99\%$ accuracy and infiltrated injection data (5.42% data) with $\approx 87\%$ accuracy with an overall accuracy of $\approx 98\%$. **Conclusion:** The objective of this work was to examine the feasibility of implementing a DL approach to PET dose injection quality monitoring.

I. INTRODUCTION

Positron Emission Tomography (PET) using 18-fluoro-2-deoxy-D-glucose (FDG) is widely used for disease staging and evaluate treatment response. The standardized uptake value (SUV) is a semi-quantitative parameter used to diagnostically assess images and perform evaluation of treatment strategies. SUV is a measure of normalization of the activity concentrations in a volume of interest (VOI) determined from PET images, and normalized by decay-corrected injected dose and the patient's body weight. The dependence of the SUV calculation on injected dose means that injection infiltration (retention of injected dose at the site of injection) can impact calculation of SUVs. This results in overestimation of the true injected dose and underestimation of the SUV which may impact the evaluation of treatment response. Dose infiltration also limits the bioavailability of the radiotracer and may prevent visualization of target disease regions.

Antecubital fossa is the most commonly used injection site which usually lies outside the Field of View (FOV) of PET imaging as the arms are often raised above head to avoid artifacts [1]. In such cases, it is difficult to trace the presence of dose infiltration, and in what way the infiltration affected the PET data. Several research works have focused on identifying the extravasated dose. Osman *et al.* studied the frequency and impact of dose infiltration [2] and reported that, in 10.5% cases, infiltration was observed which led to underestimation of SUV by an average of 11.7% in

liver and 9.3% in mediastinum. In other research works, different approaches have been proposed to correct SUV measurement based on phantom experiment [3], by Monte Carlo (MC) simulation [4] etc. However, a very recent addition to this area has been the introduction of scintillation sensors (Lucerno Dynamics, LLC, Morrisville, NC, USA) for injection quality monitoring. Time Activity Curves (TACs) during the uptake period are obtained from two sensors attached to both arms of the patient. TACs generated from the sensors on the injection arm and the opposite arm (referred to as the control arm) correspond to dose activity and baseline activity respectively. TAC analysis helps to determine if residual activity is observed at the injection site or not and can provide guidance on the severity of the impact on the injection as illustrated in Fig. 1. Fig. 1 shows TACs for good and low quality injections. In an ideal case, classified as a non-infiltration case, the TAC will indicate a fast rise in counts corresponding to the injection bolus and then quickly fall back to baseline shortly after the bolus has passed (Fig. 1a). For the TACs classified as possible infiltration cases, during the whole uptake period and even at the end of the uptake period, activity remains at the injection site (Fig. 1b and 1c).

Use of these sensors have been studied in several research works to help identifying poor injection quality, especially for the cases, where injection site is out of Field-of-View (FOV). In 2016, Williams *et al.* first demonstrated the application and feasibility of these sensors to identify and characterize infiltrations [1]. Ratio of injection arm TAC area to the control arm TAC area has been used as a threshold for detection. The study was conducted on 10 different patient cases and in 4 cases, infiltration was observed from PET data. However, with the sensor based approach, they could identify 3 cases of infiltration and the possible reasons were addressed as low intensity and volume of the injection spot. In another different study [5], it has been observed that infiltration identification by TAC analysis varied from physician's review on the basis of visual inspection of static PET images. It was found that, physician identified 15 cases of infiltration whereas, analysis based on sensor data identified 22 cases of infiltration identification out of 40 cases, leading to discrepancies in the results for 7 infiltration cases. Inconsistency observed with the existing approaches thus has been the motivation of the current study. The purpose of this work was to deploy a Deep Learning (DL) approach to perform the training of a neural network (NN) based on available dataset of injections classified as non-infiltrated and those classified as infiltrated data so that the trained NN can be

used later to classify injection data. The present form of the work implements the NN and studies the feasibility of the NN to serve as a sound classification tool.

II. MATERIALS AND METHODS

A. Deep Neural Network

Convolution Neural Network (CNN), which is a widely used DL network for classification purpose, has been used in this work to perform injection classification. LeNet5 [6], AlexNet [7], VGGNet [8], GoogleNet [9], ResNet [10] are different CNNs introduced to date and have been found to perform extraordinarily well. In this literature, a simple DL classification with slightly modified LeNET5 CNN has been performed on the TACs obtained from sensors to classify different levels of infiltrations.

TABLE I: Table of No. of Neurons in Layers

Layer	No. of Neurons
Input	600
1st Conv.Layer	16×600
1st Max-Pooling Layer	16×300
2nd Conv. Layer	32×300
2nd Max-Pooling Layer	32×150
3rd Conv. Layer	64×150
3rd Max-Pooling Layer	64×75
1st Fully Connected Layer	256
2nd Fully Connected Layer	64
2nd Fully Connected Layer/Output	2

LeNet5 is a 7 layer Convolutional Neural Network (CNN) proposed by Lecun et. al [6]. The network consisted of 2 convolutional layers followed by two subsampling/pooling layers. The convolution-pooling layers are followed by 2 FC layers. In this work, a similar architecture like LeNet5 has been implemented. The original network has been designed to feed 2D image data as input whereas, in this work, the network has been designed to operate on 1D injection data. 1D CNN has previously been used in many research works, e.g. for ECG (Electrocardiogram) signal classification [11] in the field of medical research. The model is depicted in fig. 2. A brief description of all the layers and functions are as follows:

1) *Convolutional Layer*: In the network used, 1 dimensional convolution has been performed in all the convolutional layers. Bias and weight values have been initialized with a constant value 0.1 and normal function with standard deviation 0.1 respectively. Filter kernel was set to 5.

2) *Pooling Layer*: Three pooling layers after each convolutional layers have been used to merge similar features into one and allow invariance to small shifts in the previous layer [12]. 1D max pooling operation has been performed with window size and stride of 2 for all the pooling layers.

3) *Fully Connected Layer*: Three fully connected layers followed by convolutional layers have been used with similar weight and bias initialization as convolutional layers.

4) *Activation and Loss Function*: ReLU (Rectified Linear Unit) [13] activation function has been used at the end of each convolution and FC layers except for the last FC layer.

Softmax Cross-entropy with Adam Optimizer [14] has been used to train the network.

5) *Regularization*: To reduce over-fitting, a dropout regularization has been implemented in the first two FC layers with a 0.8 probability of neurons being kept.

A learning rate of 0.0001 has been found to be optimum learning rate for the training dataset. All the hyper-parameters have been tuned by means of grid searching. The size and number of filter/kernels/feature maps for each layers have been mentioned in table I.

B. Experimental Setup & Implementation

The DL network has been implemented in python with Tensorflow library [15]. Data has been labelled first and then standardized before feeding into the network. The data was split into training and test dataset. The details are included in following subsections.

1) *Preprocessing & Standardization*: The injection data has been collected from University of Tennessee Medical Center database of injection monitoring sensor readings. The data consists of TACs from injection and control arms. The absolute difference between two TACs has been considered as input to the neural network. For each injection data, the first and last five minutes of TACs were incorporated together to feed into the NN. Thus, each input data consisted of 10 minutes i.e. 600 seconds of data with a sampling time of 1 seconds. The training data was then standardized by subtracting the mean and scaling to unit variance. Mean and standard deviation from training data has been used to be perform the standardization on test data.

2) *Labeling*: The dataset was divided into two categories: infiltrated and non-infiltrated based on the scores provided by the software in numeric range from $-\infty$ to $+\infty$. In general, negative scores were considered as non-infiltrated cases whereas positive scored data were counted as small to moderate and severe infiltration. In total 650 patient data was available, out of which 618 data (95.08%) were labeled as non-infiltrated and 32 data (4.92%) as infiltrated. The dataset was further split into training and test dataset. Table II shows the no. of data in each class of each set.

TABLE II: Table of Data Statistics

Dataset	Non-Infiltrated	Infiltrated	Total
Train	436 (94.58%)	25 (5.42%)	461
Test	182 (96.3%)	7 (3.7%)	189
Overall	618 (95.08%)	32 (4.92%)	650

III. RESULTS

The training performance was evaluated by means of 3-fold cross-validation (CV) [16]. The whole dataset was divided into 3 sets of data. 1 set was used as validation/test data and remaining 2 sets of data were used as training data each time. The neural network was trained for 300 epochs. Accuracy, specificity, sensitivity, PPV(Predictive Positive

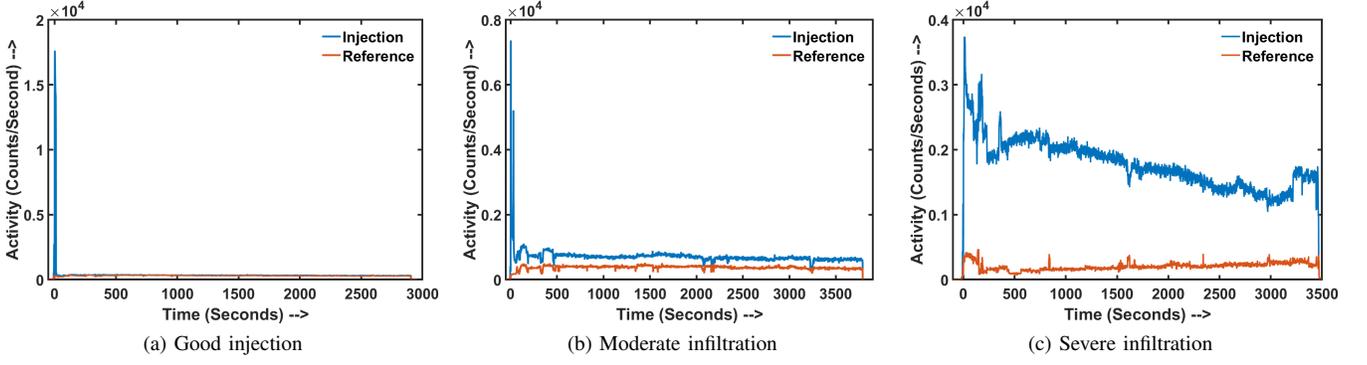


Fig. 1: Different Cases of Infiltrations: (a) Good injection quality with a fast rising uptake which fell back to the baseline shortly (b) Injection classified as a moderate infiltration where a considerable amount of activity remained at the end of the uptake period (c) Severe Infiltration where significant amount of activity remained comparing to the baseline

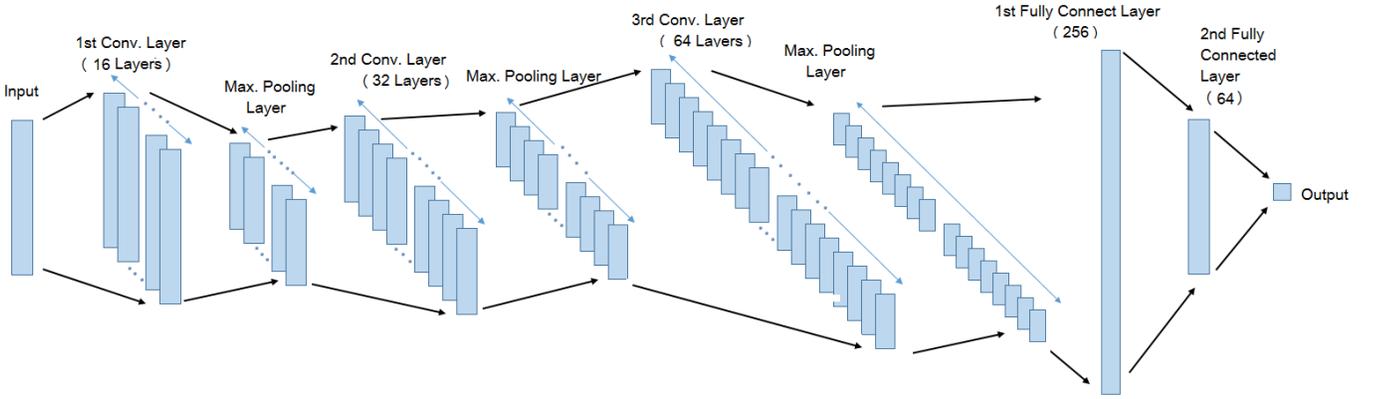


Fig. 2: Deep 1D Convolutional Neural Network used in training the TACs with 3 convolution-maxpooling layers followed by 3 fully connected layers

TABLE III: Table of Confusion Matrix & Accuracy for 3-fold Cross-Validation

Dataset	Ground Truth	Confusion Matrix		Total Data	Accuracy	Specificity	Sensitivity	PPV	NPV
		Non-infiltrated	Infiltrated						
Cross-Validation Set 1	Non-Infiltrated	137	4	141	95.45%	97.16%	76.92%	71.43%	97.86%
	Infiltrated	3	10	13					
Cross-Validation Set 2	Non-Infiltrated	146	1	147	98.70%	99.32%	85.71%	85.71%	99.32%
	Infiltrated	1	6	7					
Cross-Validation Set 3	Non-Infiltrated	148	0	148	100%	100%	100%	100%	100%
	Infiltrated	0	5	5					
Average					98.05%	98.83%	87.36%	85.71%	99.06%
Test Dataset	Non-Infiltrated	181	1	182	99.47%	99.45%	100%	87.5%	100%
	Infiltrated	0	7	7					

Value) and NPV(Predictive Negative Value) were measured as follows to perform the evaluation:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (1)$$

$$Specificity = \frac{TN}{TN + FP} \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$PPV = \frac{TP}{TP + FP} \quad (4)$$

$$NPV = \frac{TN}{TN + FN} \quad (5)$$

where, TP, TN, FP, FN denotes true positive, true negative, false positive and false negative values respectively. In this work, identifying injections classified as infiltrations has been regarded as positive and those classified as non-filtration as negative. Specificity and sensitivity provided the accuracy of identifying true negative (good quality, non-infiltration) and true positive (low quality, infiltration) respectively. PPV and NPV showed the proportion of positive and negative results. Table III includes the confusion matrices and evaluation parameters and fig. 3 and fig. 4 depicts their graphical representation.

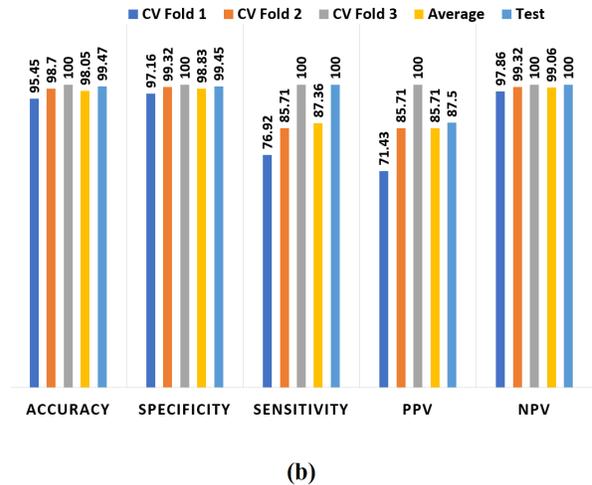
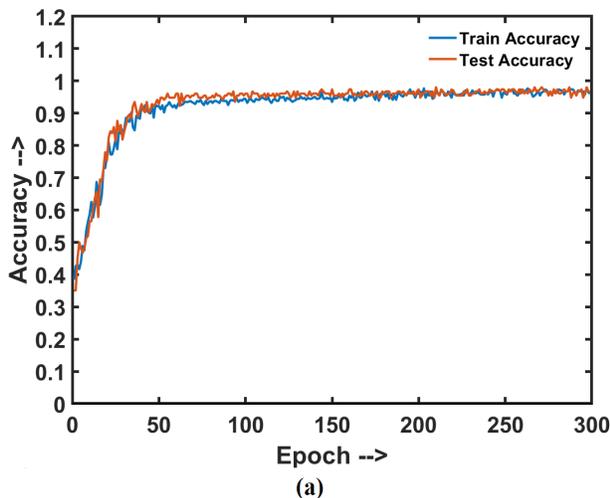


Fig. 3: (a) Train and test accuracy at each epoch during the first cross-validation iteration (b) Graphical representation of evaluation parameters for the test dataset and all three folds of cross-validation (CV Fold 1- 3) along with their average values

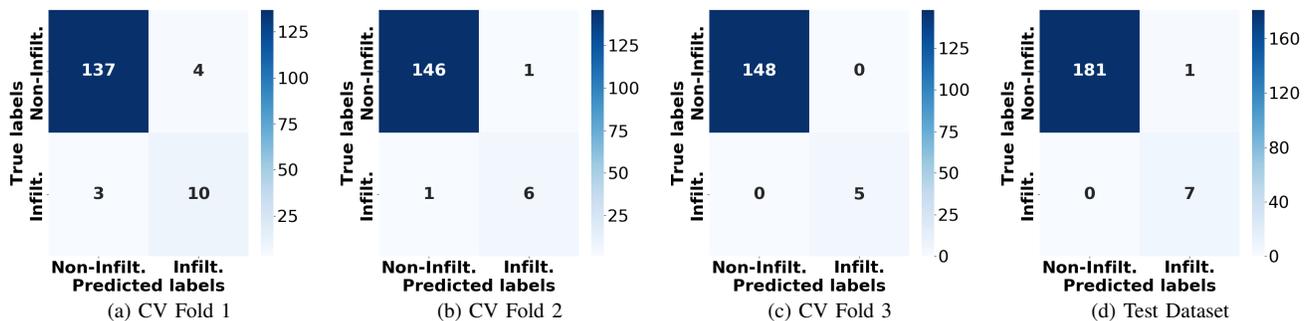


Fig. 4: Confusion Matrices for Cross-Validation fold (a) 1 (b) 2 (c) 3 and (d) Test Dataset. The colorbar on the right of each subplot indicates the number of data.

tation along with the plot of train and test accuracy. It could be seen that, for good quality injection data, average accuracy was achieved in the range of 98 – 99% (specificity) and 87% (sensitivity) for low quality injection data with overall accuracy around 98%. To the best of author’s knowledge, use of DL technique in identifying possible infiltrations is a new approach and hence standard benchmarks to compare the accuracy achieved were not readily available. Based on the research works mentioned previously, it could be seen that, in the study conducted by William *et al.*, infiltrated injections were classified with 75% accuracy, whereas in the other study by Muzaffar *et al.*, discrepancies in identifying infiltrated injections were observed for 7 out of 40 data, yielding 17.5% error i.e. 82.5% accuracy, considering one of the methods to be correct (Physician’s analysis/TAC analysis). One constraint of the current form of the work is the limited number of available data classified as infiltrated, both during training and testing. Thus, it is difficult to conclude how well the network would perform with more other low quality injection data. However, this work may interest in creating

an open access database of injection TACs which will allow performing the training and testing on a large scale of dataset and evaluate the feasibility of the implementation of NN in true clinical sense.

IV. CONCLUSION

Deep learning techniques have been found to perform extremely well in classification tasks in recent years. Use of TACs obtained from sensors to identify and classify infiltrations is also a very recent approach in PET dose infiltration research. Combination of both these techniques is a novel approach which may play a significant role in classifying infiltrated injections or identifying good injections. With this motive, a simple deep learning model has been fed with sensor data by monitoring radioactive injections. The main purpose of this study was to propose and investigate a new approach in classifying infiltration. The future work aims at training the network with higher number of infiltrated dataset and explore more sophisticated models to achieve better accuracy. Deep Convolutional Generative Adversarial Networks (DCGAN) [17] can also be implemented to

generate a large dataset and test the accuracy. Additional future work is also expected to include injection monitoring information with three-dimensional PET data to make more robust determinations of whether or not infiltration has occurred and to what degree the images may have been affected.

REFERENCES

- [1] J. M. Williams, L. R. Arlinghaus, S. D. Rani, M. D. Shone, V. G. Abramson, P. Pendyala, A. B. Chakravarthy, W. J. Gorge, J. G. Knowland, R. K. Lattanze *et al.*, "Towards real-time topical detection and characterization of fdg dose infiltration prior to pet imaging," *European journal of nuclear medicine and molecular imaging*, vol. 43, no. 13, pp. 2374–2380, 2016.
- [2] M. M. Osman, R. Muzaffar, M. E. Altinyay, and C. Teymouri, "Fdg dose extravasations in pet/ct: frequency and impact on suv measurements," *Frontiers in oncology*, vol. 1, p. 41, 2011.
- [3] K. Miyashita, N. Takahashi, T. Oka, S. Asakawa, J. Lee, K. Shizukuishi, and T. Inoue, "Suv correction for injection errors in fdg-pet examination," *Annals of nuclear medicine*, vol. 21, no. 10, pp. 607–613, 2007.
- [4] J. Silva-Rodríguez, P. Aguiar, M. Sánchez, J. Mosquera, V. Luna-Vega, J. Cortés, M. Garrido, M. Pombar, and Á. Ruibal, "Correction for fdg pet dose extravasations: Monte carlo validation and quantitative evaluation of patient studies," *Medical physics*, vol. 41, no. 5, 2014.
- [5] R. Muzaffar, S. A. Frye, A. McMunn, K. Ryan, R. Lattanze, and M. M. Osman, "Novel method to detect and characterize 18f-fdg infiltration at the injection site: A single-institution experience," *Journal of nuclear medicine technology*, vol. 45, no. 4, pp. 267–271, 2017.
- [6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [7] 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.
- [8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich *et al.*, "Going deeper with convolutions." *Cvpr*, 2015.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [11] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ecg classification by 1-d convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, 2016.
- [12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.
- [13] H. Wang and B. Raj, "A survey: time travel in deep learning space: an introduction to deep learning models and how deep learning models evolved from the initial ideas," *arXiv preprint arXiv:1510.04781*, 2015.
- [14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [15] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: <http://tensorflow.org/>
- [16] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, no. 2. Montreal, Canada, 1995, pp. 1137–1145.
- [17] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.