15th International Symposium on Computer Methods in Biomechanics and Biomedical Engineering
and
3rd Conference on Imaging and Visualization
CMBBE 2018
P. R. Fernandes and J. M. Tavares (Editors)

A MACHINE LEARNING FRAMEWORK FOR CONTEXT SPECIFIC COLLIMATION AND WORKFLOW PHASE DETECTION

Mazen Alhrishy[1], Daniel Toth[1,2], Srinivas Ananth Narayan[1,3], YingLiang Ma[4], Tanja Kurzendorfer[3], Kawal Rhode[1*], Peter Mountney[5*]

[1]School of Biomedical Engineering & Imaging Sciences, King's College London
Guys & St Thomas' NHS Foundation Trust, London, UK
mazen.m.alhrishy@kcl.ac.uk

[2]Siemens Healthineers, Frimley, UK
[3]Department of Congenital Heart Disease, Evelina London Children's Hospital, London, UK
[4]School of Computing, Electronics and Mathematics, Coventry University, Coventry, UK
[5]Siemens Healthineers, Medical Imaging Technologies, Princeton, NJ, USA
[*]Joint Senior Authors

Keywords: Machine Learning, Context Specific Collimation, Workflow Phase Detection, Image Guided Interventions, Congenital Cardiac Interventions.

Abstract: *Collimators control the field of view (FoV) by using thick blades to block X-rays leaving the source to image the patient. When the blades are adjusted to reduce the FoV, the area of the patient receiving radiation is reduced. Current fluoroscopy systems allow only for manual collimation by the operator. This can be done from the control panel using physical controls. Nonetheless, manual collimation is time consuming, causes interruption to the clinical Workflow, and is operator dependant. This is because the operator has to first identify a region of interest (RoI), then collimate around the RoI depending on the type of the procedure, workflow phase, and interventionist's preferences. In this work, we propose a learning based framework that can autonomously predict the workflow phase and localize an object of interest during congenital cardiac interventions (CCIs). In particular, we propose to learn the task of workflow recognition by using a convolutional neural network model. For training and evaluating our model, 4500 images from 25 clinical cases acquired during Biplane CCIs at Evelina London Children's Hospital, UK, were used. A training accuracy of 99% and an evaluation accuracy of 86% were achieved. The framework allows for optimal and automatic adjustment of collimation depending on the predicted workflow around the localized devices, which we refer to as context specific collimation.*

1. INTRODUCTION

Image guided interventions (IGIs) are being performed for an increasing number of procedures with longer screening time [1]. While C-arm fluoroscopy is still the imaging modality of choice for IGI, it has the disadvantage of causing radiation exposure. Children with acquired or

congenital cardiac diseases (CCD) need multiple IGIs starting from infancy, and are likely to have repetitive radiation exposure [2]. Moreover, compared with adults, children have higher radiosensitivity [3]. They also have higher proportion of the surrounding tissues exposed to radiation due to their smaller size [2].

Confining the field of view (FoV) to only that of interest reduces radiation exposure. Collimators control the FoV by using thick blades to block X-ray leaving the source to image the patient. When the blades are adjusted to reduce the FoV, the area of the patient receiving radiation is reduced. The collimator blades can be adjusted by both, the radiographer and the system. However, the X-ray system automatically adjusts collimation only in two circumstances: when digital magnification is changed (i.e. magnifying a small imaged region of interest (RoI) to cover the screen); or when the X-ray source to detector distance is changed. In both cases, the aim is to avoid unnecessary irradiation of the patient's tissue outside the area visible on the screen. The operator can manually adjust the collimation from the control panel using physical controls. When the operator adjusts the blades position, the adjusted FoV is overlaid onto the last acquired image for a preview. This means that no use of fluoroscopy is required for positioning the blades so the patient is not needlessly irradiated [4].

Nevertheless, manual collimation requires the operator to first identify the RoI, then to adjust the FoV to acquire an image of the identified RoI with optimal collimation. Adjusting the FoV depends on: 1) type of procedure, 2) the workflow phase, and 3) interventionist's preferences. For example, navigating to the target area would usually require a larger FoV, than deploying a device. A novice surgeon would also usually require a larger FoV than an experienced surgeon during the same phase to aid guidance. Optimal collimation is thus time consuming, causes interruption to the clinical workflow, and is very dependant on the operator. An automated approach is clearly beneficial. Such an approach should be able to automatically identify the RoI and to collimate around the RoI depending on the surgical workflow phase, type of procedure and devices, and surgeon preferences.

Eye controlled collimation was proposed to automatically position a dynamic collimator around a RoI the operator is gazing at. The dynamic collimator has a semitransparent plate which partially attenuates the X-rays beam and an aperture through which X-ray can leave unattenuated [5]. This, however, requires the patient or system to keep moving so that the eye-tracked RoI is kept in the unattenuated collimator region. The surgeon also has to step back from the working position into the eye tracker capture range. Moreover, during IGI procedures, surgeons has protective eyewear, thus the eye tracker might be less effective.

Another learning-based approach identifies an object of interest in an initial X-ray image then predicts the location of that object in the next X-ray image before acquisition. When the next image is acquired, the predicted location is utilized to autonomously collimate the X-ray beam to a region around that location [6]. The region size, nonetheless, is predefined, and does not depend on the workflow phase. The method was evaluated during needle path planning and needle guidance and achieved an optimal collimation compared to the initial X-ray image. The needle is a rigid high contrast object, with a straight trajectory, thus this approach might not cope with low contrast and flexible objects, such as guiding-catheters and guidewires used in IGIs.

In this paper, we propose a learning based framework that can autonomously predict the

workflow phase and identify an object of interest during congenital cardiac interventions (C-CIs). This allows the FoV to be automatically adjusted in varying amounts, depending on the predicted workflow around the identified object of interest, which we refer to as context specific collimation.

2. METHODS

2.1 Overview

The framework we propose is illustrated in Fig. 1. Once a new X-ray image is acquired, interventional devices are detected and localized in real-time. The detected workflow phase along with predefined settings can then allow for context specific collimation. The steps involved are described in more detail in the following.

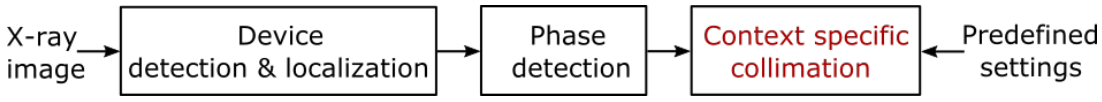

Figure 1: System overview. Present devices are detected and localized, and the workflow phase is identified. These information is then used to provide context specific collimation, together with predefined settings.

2.2 Device Detection and Localization

Real-time interventional tool extraction in IGIs is a challenging task [7, 8, 9]. This is mainly due to the low signal to noise ratio which results in low image contrast. This makes the distinction between tools and anatomical background, such as ribs and vertebrae difficult. Moreover, images are usually acquired at a high frame rate (e.g., 15 frames per second), which requires a fast detection method. Large deformation of the devices and motion artifacts are also a source of errors.

In [10], we proposed a method for detecting and localizing guiding-catheters and guidewires in CCIs. The method was developed to be used in real time by building a localized machine learning (ML) algorithm to distinguish between wires and artifacts. The potential wire-like objects were obtained from vessel enhancement filters, and input to the ML algorithm. Results showed a 83.4% success rate of detection. Detection accuracy was 0.87 ± 0.53 mm, which was measured as the error distances between the detected devices, and the manually annotated ones.

2.3 Phase Detection

Surgical workflow recognition is an active topic in the computer-assisted interventions community. Various features have been proposed for the phase recognition task, such as tool usage signals, anatomical structures, surgical actions, and visual features. A combination of these features can also be used. Several types of surgeries were investigated. These mainly included cataract, neurological, and laparoscopic surgeries [11].

To our knowledge, the task of surgical workflow recognition in IGIs has not yet been investigated. In this paper, we propose, this task can be achieved by learning the visual features from X-ray images available in the interventional room. In particular, we propose to learn the features using a convolutional neural network (CNN), which will be described in more detail in Sec. 3. The task will be, to identify the following three main phases during IGI interventions:

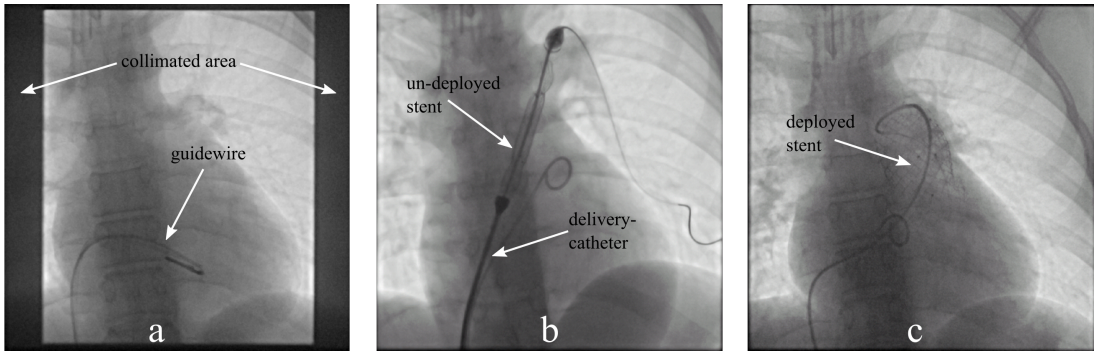

Figure 2: Example images from a clinical CCI procedure. a) During the navigation phase, b) during the pre-deployment phase, and c) during the post-deployment phase. Various interventional devices are indicated with arrows in addition to the collimated area.

- Navigation phase: During this phase, guiding-catheters and guidewires are manoeuvred from the point of access until the RoI is reached. We denote this phase as phase I. Fig. 2.a shows an example CCI image from phase I, where a guidewire can be seen.
- Pre-deployment phase: The therapeutical device (e.g., stent, balloon, valve, etc.), which is mounted on a delivery-catheter, is guided along the guidewire, and is then positioned at the targeted area, ready for deployment. We denote this phase as phase II. In Fig. 2.b, which shows an example CCI image from phase II, a delivery-catheter is present with the stent mounted and ready for deployment.
- Post-deployment phase: After the therapeutical device has been deployed, the deployment accuracy is assessed. We denote this as phase III. Fig. 2.c shows an example CCI image from phase III, where the stent has been deployed.

Figure 2 gives a concrete clinical example of supoptimal collimation. Whereas in (a) the FoV is slightly collimated (the black areas in the image), no collimation was used in (b) nor (c).

2.4 Context Specific Collimation

We define context specific collimation as the collimation corresponding to the present devices in the X-ray image and the identified phase of the workflow, together with user predefined settings. Moreover, whereas current X-ray collimators are symmetrical (i.e. vertical/horizontal blades can only be moved at the same time with the same distance), we propose using asymmetrical collimation. This can significantly reduce patient radiation dose [12], and can eliminate the

need for the RoI to be in the centre of the FoV. This would usually require constant movement of the table causing interruptions to the clinical workflow and is time consuming.

After the steps in Sec. 2.2 and Sec. 2.3 are carried out on the last acquired image, a box representing the position of asymmetrical collimator blades is shown around the localized devices as an overlay for a preview. The collimation box shows the adjusted FoV depending on the detected phase such as:

- During phase I, the FoV should show the localized devices and a wide area of the surrounding anatomy to provide enough anatomical information to aid navigation to the RoI. The size of the surrounding area will depend on the predefined settings.
- During phase II, the FoV should be collimated to include only the localized devices, as the delivery-catheter will mount the guidewire to reach the RoI.
- During phase III, ideally, the FoV should only show the deployed device and a small surrounding area to assess the accuracy of deployment. The current implementation of the method described in Sec. 2.2 can only detect and localize guiding-catheters and guidewires. Therefore, during this phase, currently, a FoV showing the localized devices and a small area of the surrounding anatomy is used. The size of the surrounding area will depend on the predefined settings.

The predefined settings offer configurable context specific collimation. For example it can include different parameters that influence the collimated area beyond the localized devices. The parameters will depend on type of procedure, type/shape of devices, level of operator's experience, expected time of imaging for aggressive/low collimation, etc.

3. EXPERIMENTS

3.1 Datasets

Images from 25 clinical cases were recorded during biplane CCI procedures at Evelina London Children's Hospital, UK. Procedures mainly included stent placement and redilation, balloon dilation, and valve placement. For training and evaluating our model, we selected 4500 images acquired from an anterior-posterior (AP) view $\pm$ 10° left/right anterior oblique (L/RAO). Images were downsampled from 512x512 pixels to 128x128 pixels, and manually annotated into three categories (i.e. labels) corresponding to the phases identified in Sec. 2.3.

Images from 20 cases were used to train the CNN. Generally, the majority of X-ray images in IGIs are acquired during phase I, thus the training dataset was balanced by discarding a significant number of the images acquired during phase I. In total, the training dataset included 3626 images. To increase the number of training images, we artificially augmented the dataset using transformations that preserve the annotated labels. These included, randomly adjusting the image brightness, contrast, and sharpness. This increased the size of our training dataset to 14504 images.

Images from the remaining five cases were used for evaluating the model prediction performance. Those included 874 images.

3.2 Architecture

The proposed model architecture is shown in Fig. 3 as represented by TensorBoard [13]. The model consists mainly of an input layer (green box), two convolutional layers (orange boxes labelled conv1 and conv2), and two fully-connected layers (gray boxes labelled fc1 and fc2).

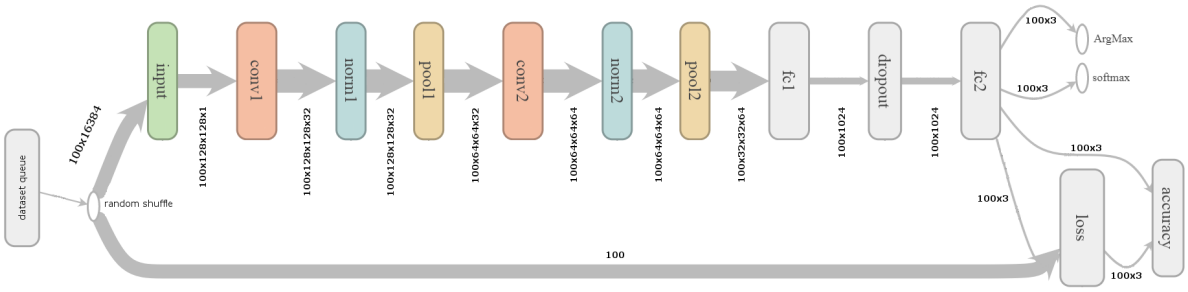

Figure 3: Model overview as represented by TensorBoard. The model consists of an input layer, two convolutional layers (conv1 and conv2), two local response normalization layers (norm1 and norm 2), two max-pooling layers (pool1 and pool2), dropout regularization layer (dropout), and two fully-connected layers (fc1 and fc2).

Input images have a size of 128x128 pixels. The first convolutional layer applies 32 filters of size 5x5 with a stride of 1 pixel. The second convolutional layer applies 64 filters of size 5x5 with a stride of 1 pixel. The output activations of both the first and second convolutional layers are followed by 1) rectified linear units (ReLU), 2) local response normalization layers, and 3) max-pooling layers with a 2x2 filter and a stride of 2 pixels.

The normalized and pooled outputs of the second convolutional layer are fed to a fully connected layer with 1024 neurons with ReLUs and dropout regularization. The second fully connected layer contains a single node for each target phase in the model, with a softmax activation function to generate a value between 0-1 for each node to represent the probability that the image falls into each target phase.

3.3 Training setup

To increase the computational efficiency, the model was trained using the Adam optimization algorithm with cross entropy loss function. The learning parameters were: learning rate=0.001, exponential decay rate for the first moment estimates=0.9, exponential decay rate for the second moment estimates=0.999. A minibatch size of 100 examples over 2000 iterations and a dropout rate after the first fully connected layer of 0.5 were used.

To provide the ReLUs with positive inputs which accelerates the early stages of learning [14], the weights in each layer were initialized from a truncated zero-mean normal distribution with a standard deviation of 0.1. The neuron biases in each layer were also initialized with the constant 0.1 for the same reason.

4. RESULTS

The model was trained for roughly 14 epochs through the training dataset of 14504 images, which took about 25 mins on an NVIDIA, Quadro K2200 GPU. Figure 4 shows the evaluation of the loss function and accuracy during training. A simple moving average smoothing using a window size of 25% of the number of points was applied to the plots for visualization (original curves are shown in faded colour). After 2000 iterations of training, an accuracy of 99% was achieved. The performance of the trained model was evaluated using the labelled 874 images from the remaining 5 cases. An evaluation accuracy of 86% was achieved.

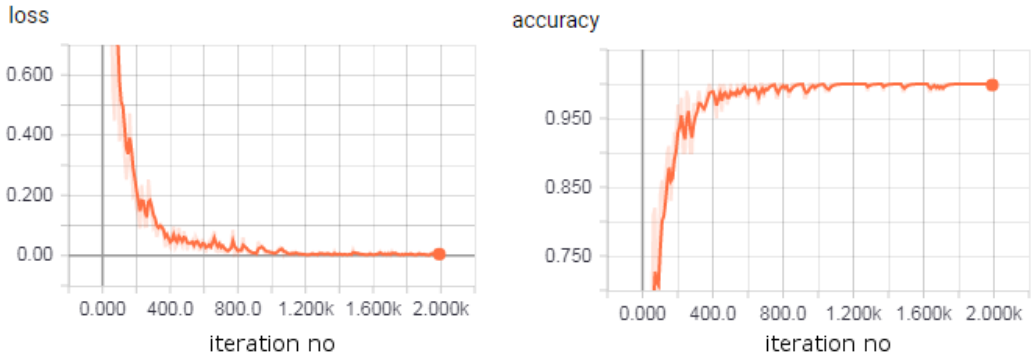

Figure 4: The loss function (left), and the training accuracy (right) plots generated during 2000 iterations of training.

Figure 5 shows the same clinical images depicted in Fig. 2. However, an example context specific collimation has been applied as follows:

- The algorithm briefly described in Sec. 2.2 was used to detect guiding-catheters and guidewires with the results overlaid onto the image in yellow.
- The workflow phase was predicted for each of the images using the trained model described in Sec. 3.2. Predicted labels were: phase I (left), phase II (middle), and phase III (right).
- Finally, context specific collimation was applied as discussed in Sec. 2.4. The size of the area beyond the detected devices (represented with arrows) was recorded in the predefined settings for a CCI, stent placement, and proficient user. These were 150 mm for phase I, and 40 mm for phase III. The area outside the collimation box was set to red to visualize the FoV outside the predicted optimal collimation.

5. CONCLUSIONS

In this paper, we have proposed a learning-based framework to provide autonomous optimal X-ray collimation during IGIs. In particular, we have trained a CNN model using X-ray images acquired during CCI procedures. The trained model can predict the surgical workflow phase to

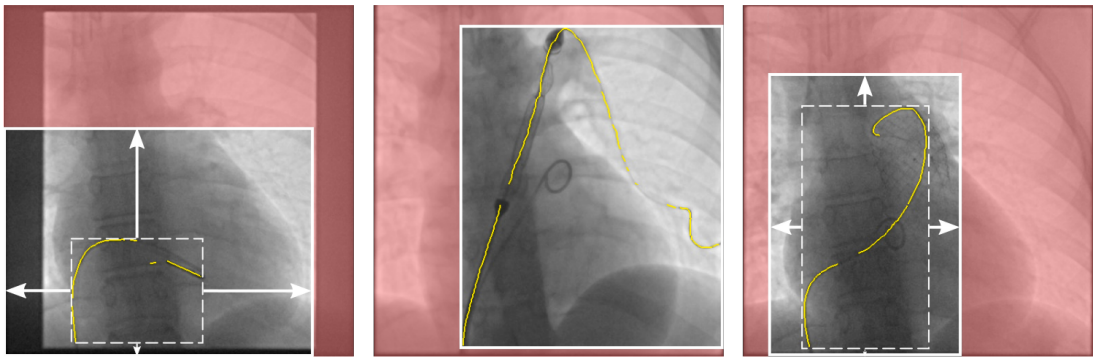

Figure 5: Example context specific collimation for the images shown in Fig. 2. Localized devices are shown in yellow. The area including only the localized devices is represented with dashed line boxes. Predicted workflow phases were: phase I, phase II, and phase III, from left to right, respectively. The predicted optimal collimation is represented with solid line boxes, with the area outside set to red.

provide context aware collimation. This is particularly important for children with CCD because of their high radiosensitivity, small size, and their need for multiple interventions starting from infancy.

The framework needs not be restricted to CCI as it can be adapted to various IGI procedures. The CNN model architecture can be adapted depending on the available dataset size and type of visual features in the acquired X-ray images. Moreover, different methods for extracting specific interventional tools and devices can be employed instead of the one presented in [10]. The task of workflow recognition can also be beneficial for automatic X-ray image indexing of various IGI procedures for training, archiving, and postoperative evaluation purposes.

Acknowledgements: This research was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. Concepts and information presented are based on research and are not commercially available. Due to regulatory reasons, the future availability cannot be guaranteed.

References

[1] C. D. Bicknell. Occupational radiation exposure and the vascular interventionalist. *Eur. J. Vasc. Endovasc. Surg.*, 2013.

[2] Mark A. Walsh, Michelle. Noga, and Jennifer. Rutledge. Cumulative radiation exposure in pediatric patients with congenital heart disease. *Pediatr. Cardiol.*, 36(2):289–294, feb 2015.

[3] H. Baysson, B. Nkoumazok, S. Barnaoui, J. L. Réhel, B. Girodon, G. Milani, Y. Boudjemline, D. Bonnet, D. Laurier, and M. O. Bernier. Follow-up of children exposed to ionising

radiation from cardiac catheterisation: the Coccinelle study. *Radiat. Prot. Dosimetry*, 165(1-4):13–16, 2015.

[4] Louis K. Wagner and Benjamin R. Archer. *Minimizing risks from fluoroscopic X rays*. Partners in Radiation Management, fifth edit edition, 2003.

[5] Stephen Balter, Dan Simon, Max Itkin, Juan F Granada, Haim Melman, and George Dangas. Significant radiation reduction in interventional fluoroscopy using a novel eye controlled movable region of interest. *Med. Phys.*, 43(3):1531–1538, 2016.

[6] Peter Mountney, Andreas Maier, Razvan Ioan Ionasec., Boese. Jan, and Comaniciu. Dorin. Method and system for obtaining a sequence of x-ray images using a reduced dose of ionizing radiation, oct 2016.

[7] Peng Wang, Terrence Chen, Ying Zhu, Wei Zhang, S. Kevin Zhou, and Dorin Comaniciu. Robust guidewire tracking in fluoroscopy. *2009 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. CVPR Work. 2009*, pages 691–698, 2009.

[8] Olivier Pauly, Hauke Heibel, and Nassir Navab. A machine learning approach for deformable guide-wire tracking in fluoroscopic sequences. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 6363 LNCS(PART 3):343–350, 2010.

[9] Hauke Heibel, Ben Glocker, Martin Groher, Marcus Pfister, and Nassir Navab. Interventional tool tracking using discrete optimization. *IEEE Trans. Med. Imaging*, 32(3):544–555, 2013.

[10] YingLiang Ma, Mazen Alhrishy, Maria Panayiotou, Srinivas Ananth Narayan, Ansab Fazili, Peter Mountney, and Kawal S. Rhode. Real-time guiding catheter and guidewire detection for congenital cardiovascular interventions. In *Int. Conf. Funct. Imaging Model. Hear.*, pages 172–182. Springer, Cham, jun 2017.

[11] A. P. Twinanda and S. Shehata and D. Mutter and J. Marescaux and M. de Mathelin and N. Padoy. EndoNet: a deep architecture for recognition tasks on laparoscopic videos. *IEEE Trans. Med. Imaging*, 36(1):86–97, 2017.

[12] Stijn De Buck, Andre La Gerche, Joris Ector, Jean Yves Wielandts, Pieter Koopman, Christophe Garweg, Dieter Nuyens, and Hein Heidbuchel. Asymmetric collimation can significantly reduce patient radiation dose during pulmonary vein isolation. *Europace*, 14(3):437–444, 2012.

[13] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, Xiaoqiang Zheng, and Google Brain. TensorFlow: a system for large-scale machine learning. *12th USENIX Symp. Oper. Syst. Des. Implement. (OSDI '16)*, pages 265–284, 2016.

[14] Alex Krizhevsky, Ilya Sutskever, and Hinton Geoffrey E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst. 25*, pages 1097—-1105, 2012.