



Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	Convolutional neural network implementation for eye-gaze estimation on low-quality consumer imaging systems
Author(s)	Lemley, Joseph; Kar, Anuradha; Drimbarean, Alexandru; Corcoran, Peter
Publication Date	2019-02-15
Publication Information	Lemley, J., Kar, A., Drimbarean, A., & Corcoran, P. (2019). Convolutional Neural Network Implementation for Eye-Gaze Estimation on Low-Quality Consumer Imaging Systems. IEEE Transactions on Consumer Electronics, 65(2), 179-187. doi: 10.1109/TCE.2019.2899869
Publisher	IEEE
Link to publisher's version	10.1109/TCE.2019.2899869
Item record	<a href="http://hdl.handle.net/10379/15473">http://hdl.handle.net/10379/15473</a>
DOI	<a href="http://dx.doi.org/10.1109/TCE.2019.2899869">http://dx.doi.org/10.1109/TCE.2019.2899869</a>

Downloaded 2024-04-23T20:20:39Z

Some rights reserved. For more information, please see the item record link above.



# Efficient CNN Implementation for Eye-Gaze Estimation on Low-Power/Low-Quality Consumer Imaging Systems

Joseph Lemley, *Student Member, IEEE*, Anuradha Kar, *Student Member, IEEE*, Alexandru Drimborean, *Member, IEEE*, and Peter Corcoran, *Fellow, IEEE*

**Abstract**—Accurate and efficient eye gaze estimation is important for emerging consumer electronic systems such as driver monitoring systems and novel user interfaces. Such systems are required to operate reliably in difficult, unconstrained environments with low power consumption and at minimal cost. In this paper a new hardware friendly, convolutional neural network model with minimal computational requirements is introduced and assessed for efficient appearance-based gaze estimation. The model is tested and compared against existing appearance based CNN approaches, achieving better eye gaze accuracy with significantly fewer computational requirements. A brief updated literature review is also provided.

**Index Terms**—Eye gaze, Neural Networks, Deep Learning

## I. INTRODUCTION

The potential of eye gaze tracking and gaze-based human computer interactions in modern consumer devices is currently an active topic for exploration. Eye gaze has been used to derive human behavioral cues, as an input modality and for achieving immersive user experiences in virtual and augmented reality systems. However, applications of gaze in consumer devices operating in real world conditions face tough challenges in terms of accuracy and reliability

### A. Gaze Tracking in Consumer Devices

After decades of research on desktop-based gaze estimation techniques, the focus has recently shifted to building eye gaze applications for dynamic platforms such as driver monitoring systems [1] and handheld devices [2]. For an automobile driver, eye based cues such as levels of gaze variation, speed of eyelid movements and eye closure can be indicative of a driver's cognitive state. These can be useful inputs for intelligent vehicles to understand driver attentiveness levels, lane change intent, and vehicle control in the presence of obstacles to avoid accidents [3].

Handheld devices like smartphones and tablets form unique platforms for gaze tracking applications wherein gaze may be used as an input modality for device control, activating safety features and novel UI designs [4]. The most challenging aspect of these modern gaze applications includes operation under

dynamic user conditions and unconstrained environments. Further requirements for implementing a consumer-grade gaze-tracking system include real-time high-accuracy operation, minimal or no calibration, and robustness to user head movements and varied lighting conditions. Therefore accurate and reliable gaze tracking typically demands high quality cameras and special equipment like narrow angle lenses, external illumination, and stereo setups [5] for capturing eye region features with sufficient details. As a result, gaze estimation systems frequently become costly with complicated setups, which are unsuitable for generic and consumer applications.

Therefore a major challenge of gaze based consumer electronics design involves maximizing system performance while reducing costs and system complexities.

### B. Deep Learning for Eye Gaze

In this paper, we introduce a calibration-free method for appearance-based gaze estimation that is suitable for consumer applications and low cost hardware with real time requirements, using a Convolutional Neural Network (CNN).

Convolutional Neural networks (CNNs) were popularized by Lecun et al. [6], who used them successfully for handwritten digit classification. These networks are inspired by the organization of the visual cortex and allow spatial information to be more efficiently learned. Convolutional Neural Networks can be used on input with any number of dimensions, but due to their success in pictures, are most popularly implemented for 2D input plus color channels. Other popular types of CNN's include 1D CNNs, which are commonly used for time series, and 3D CNNs, which can be used for volumetric data or time series data where the third dimension represents either spatial frames or temporal frames [7]. Although CNNs have become ubiquitous for most computer vision tasks, they have yet to become popular for eye gaze estimation.

### C. Contributions of this Work

From the perspective of developing a deep learning model for gaze estimation, the task can either be considered as a regression task or a classification task. Although both are useful, regression provides the greatest predictive flexibility and thus this paper treats the eye gaze estimation task as a regression problem with the goal of finding a gaze angle  $(\phi, \theta)$  that corresponds with a low resolution eye image such as one

taken from a distance with a simple RGB webcam mounted on a dashboard.

In this paper, a hardware optimized network is implemented with demonstrated suitability for deployment on such consumer devices in terms of memory requirements and speed. This network achieves superior accuracy using a dual channel input technique when compared to other state-of-the-art CNN-based gaze tracking methods for unconstrained, low resolution eye tracking.

## II. RELATED WORK

In this section a review of conventional gaze tracking techniques, studies on using low resolution data, and the application of deep learning in gaze estimation are discussed. The development and usage of important databases for gaze research are also presented.

### A. Contemporary Methods for Eye Gaze Estimation

Gaze-tracking algorithms can be broadly classified into two types: model-based methods and appearance-based methods. [8] Appearance-based methods operate directly on the eye images.

Examples of model-based methods include 2D and 3D models that use NIR illumination to create corneal reflections and track them with respect to the pupil center to estimate the gaze vector. These require polynomial or geometric approximations of the human eye to obtain the gaze direction or the point of gaze. Appearance-based methods use eye region images to extract content information such as local features, shape, and texture of eye regions to estimate gaze direction. Some key works in each of these classes of methods are summarized below.

1) *Measures of Accuracy for Eye Gaze Tracking Methods:* Contemporary research on gaze tracking measures accuracy in a wide variety of ways [9]. For example, commonly used measures include angular resolution in degrees [10], gaze recognition rates in percentage [11], and shifts in number of pixels or distance in cm/mm between gaze [12] and target locations. Unfortunately, these 4 methods are not correlated and not inter-comparable. It is the view of the authors that angular resolution is most reliable as it describes the performance of an algorithm irrespective of other system variables like user distance from tracker, pixel size of screen etc. Therefore, in this work, the angular resolution in degrees is estimated and used as the metric of accuracy for the proposed algorithm. Results from this work are only directly compared to other papers which express their performance in angles.

2) *2D Models:* 2D models utilize polynomial transformation functions for mapping the gaze vector (vector between pupil center and corneal glint) to corresponding gaze coordinates on the screen. In [13], it is shown that calibration targets and components of the mapping function are significant in determining overall accuracy of a regression-based tracker. Artificial Neural Network (ANN) based mapping methods are presented in [14], [15], [16]. In [16] a three layer ANN achieves better accuracy than regression-based approaches. Methods that are robust to head pose are presented in [17]

and [18]. Multiple geometrical transformation based mapping for handling variable user distances and head motion are used in [17]. A highly accurate calibration-free algorithm with tolerance for natural head movements is discussed in [18] using a support vector machine (SVM). Another high resolution, head-pose-invariant tracking method that does not require geometric models is presented by [19]. The typical accuracy of such models is between two and four degrees.

3) *3D Models:* 3D model-based methods typically use a geometrical model of the human eye to estimate the center of the cornea, and the optical and visual axes of the eye. Gaze coordinates are estimated as points of intersection of the visual axes with the scene. These methods achieve high accuracy (1 degree) but require elaborate system setups and knowledge about geometric relations between system components like LEDs, monitors and cameras. [20] presents a mathematical model to estimate the optical and visual axes of the users' eyes from the center of the pupil and glint, considering single and multiple cameras and light sources. Methods achieving high accuracy and head pose robustness are reported in [5], [21], [22], [23] which require multiple cameras in their setup. [22] also uses a dynamic head compensation model for updating the mapping function to track gaze under natural head movement. Calibration-free gaze estimation techniques are proposed by [24], [25] in which cameras, light sources and a spherical model of the cornea are used. Recent developments in 3D gaze tracking include usage of depth sensors along with RGB cameras such as proposed in [26]. In this, 3D gaze coordinates can be tracked in real time with a Kinect device, which provides 3D coordinates of eye features while eye parameters like eyeball and pupil center are derived from user calibration. The Kinect sensor is used in [27] for gaze estimation using free head motion along with the iris center localization method and geometric constraints-based eyeball center estimation.

4) *Appearance-Based Methods:* Appearance based methods utilize cropped eye images of a subject gazing at known locations to generate gaze point coordinates. The eye images are then used as training data for various machine learning models. For example, in [28], coordinates of eye contours, iris size, location, and pupil positions are estimated using an Active Appearance Model (AAM) and then used as input to a support vector machine (SVM) for gaze estimation. In [12], texture features are obtained using Local-Binary-Pattern (LBP) and used with an SVM along with space coordinates of the eyes for head pose free gaze tracking. A comparative evaluation of different classification methods, such as SVM, neural networks, and k-Nearest Neighbor (k-NN)s is presented in [29] where Local Binary Patterns Histograms (LBPH) and Principle Component Analysis (PCA) are used to extract eye appearance features. The use of Haar features is reported in [30], [31] for real time gaze tracking. In [32] a neural network with a skin colour model to detect face and eye regions is used. Head-pose tolerant tracking is achieved in [33] using a neural network. Recently, appearance-based methods implemented using deep learning (DL) and convolutional neural network (CNN) approaches have gained momentum. These are described in detail in section C.

## B. Gaze Estimation from Low Resolution Images

To facilitate gaze tracking in everyday settings, the use of cheap, compact and easy-to-integrate webcams is commonly preferred. Unfortunately, webcams offer low resolution images (typically 640x480 pixels) resulting in very poor gaze estimation accuracy. Low resolution images have strong noise effects [34], and distortions in the eye region contours and eye features become indistinguishable under varying illumination levels, user distance, and movements. Therefore, several approaches have been developed to achieve high gaze accuracy from low quality images and are discussed in this subsection.

An early ANN approach was used to map gaze coordinates to low quality cropped eye images in [33]. The ANN used back propagation and 50 output units each for X and Y coordinate. It was trained with 2000 image/gaze position pairs. A hybrid approach is adopted in [35] in which the iris centres are determined first using circular a Hough transform, followed by refinement using a gradient-aware random sample consensus (RANSAC) algorithm and ellipse fitting. Eye corners are estimated using Gabor jets [36] and tracked using optical flow with normalized cross-correlation. Finally, the point of gaze (POG) is estimated from the iris center and eye corners using regression. A similar method is proposed in [34], where the problem arising due to the small size of cropped eye regions from low resolution images is overcome using 2D bilinear interpolation for reconstructing the eye image to a larger size for accurate tracing of the corneal reflection vector.

In [37], multiple miniature low resolution cameras positioned around a head-mounted setup are used. The gaze-mapping function is learned from multiple cameras using a 512 unit ANN, and trained on a large dataset of eye images. In [38], the eyeball and its movement direction are detected using a deformable angular integral search (DAISMI) method followed by a Deformable template-based 2D gaze estimation (DTBGE) algorithm used as a noise filter. [39] trains an appearance model using Singular Value Decomposition (SVD) and a set of eye region images expanded artificially by adding positioning errors. Then a third order tensor is estimated from the training images as the gaze direction and positioning vectors of these images. The SVD is trained and tested by extracting the gaze vector from the test images and comparing with that obtained from the training images.

## C. Eye Gaze Estimation Using CNN's

Deep learning (DL) techniques have been successfully used in challenging conditions such as those with variable illumination, unconstrained backgrounds and free head motion. For example, [40] describes a calibration-free real-time CNN-based framework for gaze classification. Two CNNs, for the left and right eyes, are then trained independently to classify the gaze in seven directions. In [41], CNN-based gaze tracker for Augmented and Virtual reality devices achieves foveated rendering and gaze-contingent focus. The deep learning model is built to be robust to variations like skin and eye colour, illumination and occlusion. In [42], deep features are obtained from eye images using multi-scale convolutions and pooling for predicting gaze direction. This method uses minimized

cross-entropy loss, coupled with Random Forest regression as a clustering algorithm. It classifies areas on a device screen according to gaze locations, and operates under natural illumination and head poses. [43] describes a novel, appearance-based gaze estimation method in which a CNN utilizes the full face image as input with spatial weights on the feature maps to suppress or enhance information in different facial regions. It achieves high accuracy and robust performance under varied illumination and extreme head poses. [44] achieves free head pose, 3D gaze tracking using two separate head pose and eye movement models with two CNNs, connected via a gaze transform layer. Finally, in [45] a CNN is built to learn the mapping between 2D head angle, eye image and gaze angle (output) using a small Lenet-inspired CNN. For testing, an extensive database is built with more than 200,000 images under variable illumination levels and eye appearances. This database (called MPII Gaze) is also used in this work and its further details are provided in the next section.

Several of the CNN-based works are specifically targeted towards gaze tracking in consumer/handheld devices such as [46], [47]. In [46], a CNN-based real time, calibration-free gaze estimation algorithm is presented. It is trained using a large and diverse dataset of eye images taken under variable lighting, head pose, and backgrounds captured from users through a smartphone app. Inputs to their CNN model include eye and face images. The location of the faces in the images are obtained through a face grid, which is used to infer relative eye and head poses. [47] presents a calibration-free method using Deep Belief Networks which classify gaze into a grid of nine gaze locations under various head-poses and viewing directions. In [48], a nine directional CNN-based gaze classifier is developed for a screen typing application, robust to false detections, blinks, and saccades (rapid, abrupt changes in fixation).

An overview of selected deep learning methods for gaze estimation can be seen in table I.

## D. Related Work Utilizing the MPII Gaze Dataset

The MPII Gaze dataset[45] is large and challenging, containing images collected under a wide range of realistic scenarios, such as varied illumination levels, eye appearances and head poses. Use of MPII gaze dataset for training and testing gaze estimation algorithms can be found in their own paper [45], as also in [43], [49], [57], [58]. [45], which introduces the MPII Gaze dataset, uses a multimodal CNN for gaze estimation and reports a cross dataset test accuracy of 6 degrees. [43] uses full face (instead of eye only and multi-region) images with a CNN and achieves a person independent accuracy of 4.8 degrees on MPII Gaze while being robust to illumination variations and extreme head poses. In [49], gaze location over a block of screen area is tracked using a CNN, and the cross-subject performance is tested with MPII Gaze and that authors' own dataset. On MPII Gaze, the classification accuracy is poor (75.6%) compared to that on authors' dataset (92.5%). A Deep Regression Bayesian Network described in [58] achieves an accuracy of 7.1 degrees when tested on this dataset. In [59], [60], [61], the MPII Gaze dataset is used for

TABLE I  
SELECTED CNN MODELS FOR EYE GAZE ESTIMATION

Citation	Dataset used	Network model used	Image Resolution	Accuracy	Special features
[41]	CAVE and Own dataset (cropped images of the eye and their respective gaze pixel coordinates)	LeNet (two convolutional layers and two pooling layers followed by a fully connected layer at the end). 21 classes for CAVE , 1829 classes for captured dataset.	28x28	6.7 degrees tested on CAVE as well as own collected dataset	Near eye tracking, operating under lighting changes and occlusions
[43]	MPII Gaze, UT Multiview	AlexNet. five convolutional layers, two fully connected layers. Additional linear regression layer on top of last fully connected layer.	448 x 448 (full face)	4.8 degrees on MPII Gaze, 6 degrees on UT dataset. Person-independent evaluation	Tolerates various illumination, gaze direction,
[42]	Own dataset with 107,681 images under different lighting, head movement, glasses.	Three convolutional layers with max-pooling layer, one single max-pooling layer, one hidden layer and one soft-max layer	40x70	5-7 degrees within dataset evaluation, but training and test set have different gaze angles.	Works under natural light with free head motion
[44]	Own dataset 200-subjects different head poses, eyeball movements, lighting conditions, glasses, occlusions, reflections.	AlexNet, BN-Inception network. Two CNNs to model head pose and eyeball motion. A gaze transform layer to aggregate them into gaze prediction	62x62	4.3 degrees cross subject evaluation	Allows free head motion
[45]	MPII Gaze, Eyediap, UT Multiview	LeNet architecture. A linear regression layer trained on top of fully connected layer. Multimodal CNN model to use eye image and head pose information	60x36	~6 degrees Cross Dataset Evaluation	Variable appearance, illumination, head pose
[46]	Gaze Capture dataset	AlexNet + SVR	80x80	2.58cm (within dataset evaluation)	Tolerant pose, appearance, and lighting
[49]	Own dataset. 56 groups of eye videos, 181440 eye images from 22 subjects	3 convolutional layers followed by max-pooling layers, 2 fully connected and a soft-max layer for classification. 6 & 54 classes	40x72	6.375 deg, 69.3% Cross dataset evaluation	---

comparing synthetic datasets and facial models. [59] presents a method for synthesizing a large set of variable eye region images with a generative 3D eye region model. Then a gaze estimation method using the k-Nearest-Neighbour algorithm) is tested on the synthetic data and the MPII Gaze dataset to achieve an accuracy of 9.95 and 9.58 degrees respectively. Another method for synthetic, labelled photo-realistic eye region image creation is described in [60] using head scan geometry. The generated dataset, along with MPII gaze, is used to test and compare the accuracy of a CNN based gaze estimation method. In [61], a facial behaviour analysis tool is developed that is capable of tracking gaze vectors using a Conditional Local Neural Fields (CLNF) framework by detecting eye region features like eyelids, irises and pupils. The tool is tested on MPII Gaze to achieve 9.96 degrees of accuracy in gaze estimation. In [62] a semi-supervised learning method is developed for improving the realism of simulated data and used to create refined training images for gaze estimation using CNNs. The CNNs are then tested on the MPII gaze dataset to achieve an error of 7.8 degrees. Apart from the above, the MPII dataset has been used for training a CNN for gaze estimation coupled with gaze target discovery in [63] and a cascaded-regressor-based eye center detector [64].

### III. IMAGE AND VIDEO DATASETS FOR LOW RESOLUTION AND UNCONSTRAINED GAZE ESTIMATION

A survey of relevant publicly available gaze databases is summarized in Table II. In this survey, only the databases built for training and testing gaze estimation algorithms are listed,

while other gaze databases, e.g. for studying saliency models, are not included as they are out of scope for this work.

## IV. METHODS

The CNN-based gaze estimation methods in this work were evaluated on NVIDIA 1080 TI GPUs using python 2.7 and caffe 1.0 with accuracy and euclidean loss layers modified to calculate angle difference in radians. Person-exclusive, leave-one-out cross-validation was used in all experiments.

In eye gaze tracking literature it is common to use the word “accuracy” and “error” interchangeably and this can sometimes cause confusion to the reader. For this reason we use the word “error” in any case where the meaning could be unclear. All angles are reported in degrees. In this paper, error was determined as the average euclidean distance between the ground truth and predicted angles on the left-out, person-exclusive test set.

Multiple deep neural networks were compared for eye gaze estimation using deep neural networks. The publicly available MPII Gaze dataset was used for all experiments except for the first where the UT Multiview dataset is also used.

### A. MPII Gaze Dataset Details

The MPII gaze dataset is a large collection of 213659 images captured under unconstrained conditions from 15 subjects over several days. The images are collected under multiple illumination conditions. Some of the subjects wear spectacles and some do not. The images were captured at various gaze

TABLE II  
PUBLICLY AVAILABLE DATASETS FOR GAZE ESTIMATION

Name	Persons	Items	Conditions	Resolution	Purpose
MPII Gaze [45]	15	213,659 images.	Software running on subjects' laptops ask participants to look at a random 20 on-screen positions and confirm	unknown	Appearance-based gaze estimation in the wild.
UT Multiview [50]	50	64000 eye images	160 gaze directions per person were acquired using 8 cameras (views)	SXGA resolution (1280x1024)	Training and test data for appearance-based gaze estimation methods.
EYEDIAP [51]	16	94 sessions	Diversity of participants, head poses, gaze targets and sensing conditions. Screen or 3D objects. Data collection with Kinect for RGB and depth video streams	640x480 at 30 fps	Training and evaluation of gaze estimation approaches with robustness to pose, person.
Gaze Capture [46]	1474	2445504 images	Data captured with iphone/ipad using app. Large variation in pose, appearance, lighting. Variation in relative distance and orientation of the mobile device	unknown	Training CNNs for high accuracy calibration-free eye tracking on handheld devices under variable conditions.
TabletGaze [52]	51	100000 images	Video sequences recorded with tablet front-facing camera while subjects look at a dot on tablet screen. Unrestricted subject motion, each subject performed 4 body postures: standing, sitting, slouching, and lying.	1280x720	Mobile gaze dataset for studying unconstrained mobile gaze estimation
Weidenbacher [53]	20	2220 images	Manual landmarks on pupils, eye corners, nose tip, mouth corners. Horizontal head rotations (0 to 90 in steps of 10 ), vertical head poses 0, 30, and 60 azimuth (-20, +20) elevation. For each head pose, nine different gaze conditions	1600x1200	Evaluating computational methods for head pose and eye gaze estimation
McMurrough [54]	20	120 sessions	Videos recorded eye motion as subjects look at, or follow a set of predefined points on a computer screen. Head position in 3D captured using a Vicon Motion Tracking System	768x480 pixels at a frame rate of 29.97 Hz	To be used as a benchmark for Point of Gaze (PoG) detection algorithms
OMEG [55]	50	40000 images	Eye images captured under multiple head poses, three fixed poses, 0 and 30 degree , and a free pose style.	1280 x 1024	Evaluating and comparing gaze tracking algorithms
HPEG [56]	10	20 videos	Subject faces camera frontally, free to move, background covers a big part of the image, with intense human action.	1280x960 pixels, 30 fps	Head pose and gaze estimation algorithm testing

angles, recorded by software running on the participant's laptops. In each session, the subjects were asked to look at random sequences of 20 onscreen positions and to confirm their attentiveness, the subjects were asked to press the space bar once the onscreen target was disappearing.

The dataset contains eye and head features and target (gaze angle) values for every participant. To use MPII Gaze, the authors suggest mapping their reported vector to angles using a Rodrigous transformation, and this has been done for all reported experiments.

## V. EXPERIMENTS AND RESULTS

In this section, multiple experiments are described to provide insight on 4 primary research questions. These are tested on multiple CNN architectures and are discussed in this section. One of the first research goals was to achieve state of the art test error on a network that could perform inference within 3-15 ms on a typical single proprietary low power consumer embedded device.

The specific research questions are:

- 1) How does an architecture that uses both eyes compare to one that uses one eye in terms of accuracy?
- 2) How does simulated camera distance impact eye gaze accuracy for the proposed model?

3) Can augmentation be used to reduce any negative impacts?

4) Can the proposed hardware-friendly architecture perform with sufficient accuracy and speed?

First, the intra-dataset, person-exclusive experiments from [45] were duplicated. The same procedure to estimate accuracy was used except the altered accuracy layers were modified to eliminate NaNs by replacing undefined values of the arc cosine function with the largest or smallest valid values as appropriate.

### A. Approach 1: Analysis of Eye Flipping

In [45], one network is used for both eyes and one of the eyes is flipped so that the gaze angle is roughly correct. An experiment was designed to see if this flipping had an impact on model accuracy. Six experiments were performed using the UT and MPII-Gaze datasets to see if training on both eyes or just one eye impacted accuracy. By doing experiments with combined and non combined datasets, it was also possible to determine if they had similar distributions, and thus, if combining the two would be helpful for future experiments.

As shown in table III, the method of individually classifying eye images and simply adjusting the right eye and angles as used in [45] is a limiting factor in accuracy for that method. In both datasets, the performance was increased exclusively using

TABLE III  
RESULTS OF APPROACH 1

Training set	Error
MPII Left eye only	5.9 degrees
MPII+UT Left eye only	5.6 degrees
MPII Both eyes	7.4 degrees
MPII+UT Both eyes	6.5 degrees
MPII Right eye only	5.3 degrees
MPII+UT Right eye only	6.1 degrees

left eyes or right eyes. This suggests that simply flipping the eye as suggested by [45] may be a source of error in their model.

These results also indicated that the distribution of MPII Gaze and UT-Multiview are sufficiently different that combining the two for training gives no or very little improvement. Because of this, it was decided to use only MPII Gaze for the remaining experiments in this section as UT-Multiview had no significant influence on error.

### B. A New Approach: Dual Eye Channels

Given the problems identified in the previous subsection with flipping one of the eyes, and not wanting to use two different networks for reasons of efficiency, a new approach involving using both the left and right eyes in separate input channels was investigated. Specifically, the left eye and right eye images are passed to the network in channels 0 and 1 respectively, and the gaze and pose information are averaged between the left and right eye images to create a single gaze and pose vector. Due to the results in the previous section, which indicated that data from UT did not significantly impact the results, only the MPII-Gaze dataset was used.

This modified, two channel architecture resulted in a significant increase in accuracy, averaging 4.63 degrees of error between the target and predicted values on unseen individuals. A diagram of this network can be seen in Figure 1.

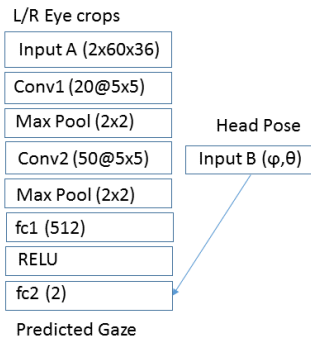


Fig. 1. Diagram of hardware CNN for initial eye gaze estimation tasks, heavily based on network from [45]

### C. Can We Reduce the Number of Parameters?

Deep neural networks can often be made more efficient by reducing the number of parameters but this can sometimes come at the cost of accuracy. To see if reducing the number of parameters was possible without harming accuracy, an

TABLE IV  
RESULT OF DISTANCE SIMULATION EXPERIMENTS

Resolution	Error
60 x 36	4.63 degrees
52 x 31	9.90 degrees
26 x 16	10.10 degrees

experiment was performed to halve the size of all output parameters. This experiment was not allowed to run for the full duration because the exact angle accuracy did not matter, only evidence that the network complexity could be reduced to a point where it would be small enough if necessary. This resulted in an average error of 4.980% on an unseen individual from the MPII Gaze dataset and indicates that reducing the number of parameters had little impact on accuracy.

### D. Multi Resolution Experiments

Eye gaze systems in consumer devices must be able to maintain accuracy at a large range of distances. Although MPII Gaze has some variability in distance from the camera, the distances are not realistic for the conditions expected in, for example, a driver monitoring system or a distant cell phone camera. Specifically, it was desired to accommodate realistic distances between the camera and the subject in situations that would be typical in commercial eye trackers that utilize low cost, low resolution cameras. To simulate the loss of information caused by distance, down-sampling was performed on the eye images in MPII-Gaze as follows:

- Input image 60 x 36 ->Downscale to 52 x 31 ->Upscale to 60 x 36 ->CNN Eye gaze angle
- Item Input image 60 x 36 ->Downscale to 26 x 16 ->Upscale to 60 x 36 ->CNN Eye gaze angle.

As can be seen in table IV, the network learned a narrow range of distances, and performance deteriorates when the subject is further from the camera than those in the training set. As a sanity check, an experiment was done to see if the Downscaling algorithm was at fault for the poor results, so in addition to nearest, we also tried bicubic, linear, and LANCZOS from using OpenCV. The experiment showed that the downscaling algorithm used had no influence on the results.

This demonstrates that the model is sensitive to changes in distance. In the next section, an experiment is performed to see if data augmentation can be used to improve upon this.

### E. Impact of Random Resizing as Augmentation

Data augmentation has been shown in many studies to have a large impact on model performance.

To further improve accuracy, the dataset was augmented with multiple randomly chosen resolutions to match the full range of desired distances. To help reduce the chance that the network would learn the specific interpolation method used, Nearest is used in the training set, but Lanczos filtering is used in the testing set.

- Original resolution: 4.918 degrees error
- 60 x 36 ->52 x 31 ->60 x 36 : 4.94 degrees error

TABLE V  
ERROR OF PROPOSED MODEL FOR VARIOUS RESOLUTIONS (DEGREES)

Resolutions	36 x 60	31 x 52	26 x 16
Best	3.650	4.1690	4.240
Second best	4.100	4.324	4.366
Benchmark	4.917	4.940	4.970

TABLE VI  
COMPARISON OF PROPOSED MODEL WITH OTHER PUBLISHED WORKS ON THE MPII-GAZE DATABASE

Citation	Error (degrees)
[61]	9.96
[59]	9.58
[62]	7.8
[58]	7.1
[45]	6
[43]	4.8
Proposed	3.64

- 60 x 36 ->26 x 16 ->60 x 36: 4.97 degrees error

These results indicate that augmenting the images with distances that are likely to be encountered in real world usage situations is an effective way to increase accuracy and succeeds in achieving some invariance to subject distance.

#### F. A Quest for Hardware Efficiency and Even Better Accuracy

It was shown in [65] that two stacked layers of 3x3 convolutions has the same receptive field as a single 5x5 layer, with fewer multiplications. Due to this, one of the requirements was that kernel sizes be 3x3, this required retraining and slight redesign of the network. Several experiments involving architectures with stacked 3x3 kernels were performed using different parameters. The best two architectures were further evaluated, and the best models from each of them were chosen and evaluated on multiple resolutions as shown in table V. A diagram of the final architecture used can be seen in figure 2, and a comparison with other published works can be seen in table VI.

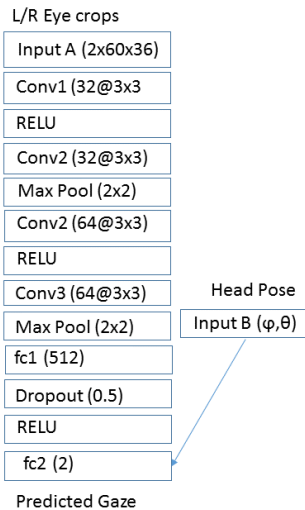


Fig. 2. Diagram of hardware optimized CNN for eye gaze estimation corresponding to the accuracies in table V.

## VI. CONCLUSION

Our results show that using information from both eyes in the neural network can increase accuracy. This is demonstrated in section V, where adding additional eye information from the opposite eye enabled improved results over individual eyes, helping the network make sense of low quality images with ambiguous gaze. As expected, in all cases, the deeper network had the best performance. This research demonstrated the sensitivity of such models to variations in distance and how data augmentation can be used to overcome this. Most importantly, a new compact hardware-friendly architecture designed for use in small consumer electronics has been introduced and evaluated on the eye gaze task.

When evaluated on MPII Gaze, the proposed model performs favorably (see table VI even when compared with much larger networks in the literature).

Since augmentation resulted in a significant improvement in accuracy, it may be fruitful to try other types of augmentation such as Generative Adversarial Networks (GANS) with landmarks, [66] and Smart Augmentation (SA) [67] in future work. This will either require modifying such methods to work on regression problems or translating the eye gaze problem into a classification task for the purpose of generating augmented data[68] and then back to a regression task. Additionally, there are plans to investigate whether temporal information[69] can be used to further increase the accuracy without sacrificing the need for performance as it has been shown to increase performance in DMS systems.

## ACKNOWLEDGMENT

This research is funded under the SFI Strategic Partnership Program by Science Foundation Ireland (SFI) and FotoNation Ltd. Project ID: 13/SPP/I2868 on Next Generation Imaging for Smartphone and Embedded Platforms. This work is also supported by an Irish Research Council Employment Based Programme Award. Project ID: EBPPG/2016/280.

## REFERENCES

- [1] Y. Liang, M. L. Reyes, and J. D. Lee, "Real-time detection of driver cognitive distraction using support vector machines," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 340–350, June 2007.
- [2] E. Wood and A. Bulling, "Eyetab: Model-based gaze estimation on unmodified tablet computers," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, ser. ETRA '14. New York, NY, USA: ACM, 2014, pp. 207–210. [Online]. Available: <http://doi.acm.org/10.1145/2578153.2578185>
- [3] A. Tawari and M. M. Trivedi, "Robust and continuous estimation of driver gaze zone by dynamic analysis of multiple face videos," in *2014 IEEE Intelligent Vehicles Symposium Proceedings*, June 2014, pp. 344–349.
- [4] V. Vaitukaitis and A. Bulling, "Eye gesture recognition on portable devices," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ser. UbiComp '12. New York, NY, USA: ACM, 2012, pp. 711–714. [Online]. Available: <http://doi.acm.org/10.1145/2370216.2370370>
- [5] S.-W. Shih and J. Liu, "A novel approach to 3-d gaze tracking using stereo cameras," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 234–245, Feb 2004.
- [6] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.



- [7] J. Lemley, S. Bazrafkan, and P. Corcoran, "Deep learning for consumer devices and services: Pushing the limits for machine learning, artificial intelligence, and computer vision." *IEEE Consumer Electronics Magazine*, vol. 6, no. 2, pp. 48–56, 2017.
- [8] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 3, pp. 478–500, 2010.
- [9] A. Kar and P. Corcoran, "A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms," *IEEE Access*, vol. 5, pp. 16495–16519, 2017.
- [10] F. L. Coutinho and C. H. Morimoto, "Augmenting the robustness of cross-ratio gaze tracking methods to head movement," in *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 2012, pp. 59–66.
- [11] S. Chen and C. Liu, "Eye detection using discriminatory haar features and a new efficient svm," *Image and Vision Computing*, vol. 33, pp. 68–77, 2015.
- [12] H. chuan Lu, C. Wang, and Y. w. Chen, "Gaze tracking by binocular vision and lbp features," in *2008 19th International Conference on Pattern Recognition*, Dec 2008, pp. 1–4.
- [13] P. Bignaut, "Mapping the pupil-glint vector to gaze coordinates in a simple video-based eye tracker," *Journal of Eye Movement Research*, vol. 7, no. 1, 2013. [Online]. Available: <https://bop.unibe.ch/index.php/JEMR/article/view/2373>
- [14] Z. Zhu and Q. Ji, "Eye and gaze tracking for interactive graphic display," *Machine Vision and Applications*, vol. 15, no. 3, pp. 139–148, Jul 2004. [Online]. Available: <https://doi.org/10.1007/s00138-004-0139-4>
- [15] C. Jian-nan, Z. Chuang, Y. Yan-tao, L. Yang, and Z. Han, "Eye gaze calculation based on nonlinear polynomial and generalized regression neural network," in *2009 Fifth International Conference on Natural Computation*, vol. 3, Aug 2009, pp. 617–623.
- [16] J. Wang, G. Zhang, and J. Shi, "2d gaze estimation based on pupil-glint vector using an artificial neural network," *Applied Sciences*, vol. 6, no. 6, 2016. [Online]. Available: <http://www.mdpi.com/2076-3417/6/6/174>
- [17] Chunfei, K.-A. Ma, B.-D. Choi, and S.-J. K. Choi, "Robust remote gaze estimation method based on multiple geometric transforms," *Optical Engineering*, vol. 54, pp. 54 – 54 – 7, 2015. [Online]. Available: <https://doi.org/10.1117/1.OE.54.8.083103>
- [18] Z. Zhu, Q. Ji, and K. P. Bennett, "Nonlinear eye gaze mapping function estimation via support vector regression," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 1, 2006, pp. 1132–1135.
- [19] J. Zhu and J. Yang, "Subpixel eye gaze tracking," in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, May 2002, pp. 124–129.
- [20] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1124–1133, June 2006.
- [21] T. Ohno and N. Mukawa, "A free-head, simple calibration, gaze tracking system that enables gaze-based interaction," in *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, ser. ETRA '04. New York, NY, USA: ACM, 2004, pp. 115–122. [Online]. Available: <http://doi.acm.org/10.1145/968363.968387>
- [22] Z. Zhu and Q. Ji, "Novel eye gaze tracking techniques under natural head movement," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 12, pp. 2246–2260, Dec 2007.
- [23] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2, June 2003, pp. II–451–8 vol.2.
- [24] T. Nagamatsu, J. Kamahara, and N. Tanaka, "Calibration-free gaze tracking using a binocular 3d eye model," in *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '09. New York, NY, USA: ACM, 2009, pp. 3613–3618. [Online]. Available: <http://doi.acm.org/10.1145/1520340.1520543>
- [25] D. Model and M. Eizenman, "User-calibration-free remote eye-gaze tracking system with extended tracking range," in *2011 24th Canadian Conference on Electrical and Computer Engineering (CCECE)*, May 2011, pp. 001268–001271.
- [26] K. Wang and Q. Ji, "Real time eye gaze tracking with kinect," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, Dec 2016, pp. 2752–2757.
- [27] X. Zhou, H. Cai, Z. Shao, H. Yu, and H. Liu, "3d eye model-based gaze estimation from a depth sensor," in *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec 2016, pp. 369–374.
- [28] Y.-L. Wu, C.-T. Yeh, W.-C. Hung, and C.-Y. Tang, "Gaze direction estimation using support vector machine with active appearance model," *Multimedia Tools and Applications*, vol. 70, no. 3, pp. 2037–2062, Jun 2014. [Online]. Available: <https://doi.org/10.1007/s11042-012-1220-z>
- [29] C. M. Yilmaz and C. Kose, "Local binary pattern histogram features for on-screen eye-gaze direction estimation and a comparison of appearance based methods," in *2016 39th International Conference on Telecommunications and Signal Processing (TSP)*, June 2016, pp. 693–696.
- [30] S. Chen and C. Liu, "Eye detection using discriminatory haar features and a new efficient svm," *Image Vision Comput.*, vol. 33, no. C, pp. 68–77, Jan. 2015. [Online]. Available: <http://dx.doi.org/10.1016/j.imavis.2014.10.007>
- [31] Y. Li, X. Xu, N. Mu, and L. Chen, "Eye-gaze tracking system by haar cascade classifier," in *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, June 2016, pp. 564–567.
- [32] T. Schneider, B. Schauerte, and R. Stiefelhagen, "Manifold alignment for person independent appearance-based gaze estimation," in *Proceedings of the 2014 22nd International Conference on Pattern Recognition*, ser. ICPR '14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 1167–1172. [Online]. Available: <http://dx.doi.org/10.1109/ICPR.2014.210>
- [33] S. Baluja and D. Pomerleau, "Non-intrusive gaze tracking using artificial neural networks," in *Advances in Neural Information Processing Systems*, 1994, pp. 753–760.
- [34] Y. Fu, W. P. Zhu, and D. Massicotte, "A gaze tracking scheme with low resolution image," in *2013 IEEE 11th International New Circuits and Systems Conference (NEWCAS)*, June 2013, pp. 1–4.
- [35] A. George and A. Routray, "Fast and accurate algorithm for eye localisation for gaze tracking in low-resolution images," *IET Computer Vision*, vol. 10, no. 7, pp. 660–669, 2016.
- [36] D. Gonzalez-Jimenez and J. L. Alba-Castro, "Shape-driven gabor jets for face description and authentication," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 4, pp. 769–780, Dec 2007.
- [37] M. Tonsen, J. Steil, Y. Sugano, and A. Bulling, "Invisibleye: Mobile eye tracking using multiple low-resolution cameras and learning-based gaze estimation," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 106:1–106:21, Sep. 2017. [Online]. Available: <http://doi.acm.org/10.1145/3130971>
- [38] Y.-T. Lin, R.-Y. Lin, Y.-C. Lin, and G. C. Lee, "Real-time eye-gaze estimation using a low-resolution webcam," *Multimedia Tools Appl.*, vol. 65, no. 3, pp. 543–568, Aug. 2013. [Online]. Available: <http://dx.doi.org/10.1007/s11042-012-1202-1>
- [39] Y. Ono, T. Okabe, and Y. Sato, "Gaze estimation from low resolution images," in *Advances in Image and Video Technology*, L.-W. Chang and W.-N. Lie, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 178–188.
- [40] A. George and A. Routray, "Real-time eye gaze direction classification using convolutional neural network," in *2016 International Conference on Signal Processing and Communications (SPCOM)*, June 2016, pp. 1–5.
- [41] R. Konrad, S. Shrestha, and P. Varma, "Near-eye display gaze tracking via convolutional neural networks."
- [42] Y. Wang, T. Shen, G. Yuan, J. Bian, and X. Fu, "Appearance-based gaze estimation using deep features and random forest regression," *Know.-Based Syst.*, vol. 110, no. C, pp. 293–301, Oct. 2016. [Online]. Available: <https://doi.org/10.1016/j.knsys.2016.07.038>
- [43] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "It's written all over your face: Full-face appearance-based gaze estimation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, July 2017, pp. 2299–2308.
- [44] H. Deng and W. Zhu, "Monocular free-head 3d gaze tracking with deep learning and geometry constraints," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 3162–3171.
- [45] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "Appearance-based gaze estimation in the wild," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 4511–4520.
- [46] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba, "Eye tracking for everyone," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 2176–2184.
- [47] H. Park and D. Kim, "Gaze classification on a mobile device by using deep belief networks," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, Nov 2015, pp. 685–689.
- [48] C. Zhang, R. Yao, and J. Cai, "Efficient eye typing with 9-direction gaze estimation," *Multimedia Tools and Applications*, Nov 2017. [Online]. Available: <https://doi.org/10.1007/s11042-017-5426-y>
- [49] X. Wu, J. Li, Q. Wu, and J. Sun, "Appearance-based gaze block estimation via cnn classification," in *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, Oct 2017, pp. 1–5.

- [50] Y. Sugano, Y. Matsushita, and Y. Sato, "Learning-by-synthesis for appearance-based 3d gaze estimation," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 1821–1828. [Online]. Available: <http://dx.doi.org/10.1109/CVPR.2014.235>
- [51] K. A. Funes Mora, F. Monay, and J.-M. Odobez, "Eyediap: A database for the development and gaze evaluation algorithms from rgb and rgb-d cameras," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, ser. ETRA '14. New York, NY, USA: ACM, 2014, pp. 255–258. [Online]. Available: <http://doi.acm.org/10.1145/2578153.2578190>
- [52] Q. Huang, A. Veeraraghavan, and A. Sabharwal, "Tabletgaze: dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets," *Machine Vision and Applications*, vol. 28, no. 5, pp. 445–461, Aug 2017. [Online]. Available: <https://doi.org/10.1007/s00138-017-0852-4>
- [53] U. Weidenbacher, G. Layher, P. M. Strauss, and H. Neumann, "A comprehensive head pose and gaze database," in *2007 3rd IET International Conference on Intelligent Environments*, Sept 2007, pp. 455–458.
- [54] C. D. McMurrough, V. Metsis, J. Rich, and F. Makedon, "An eye tracking dataset for point of gaze detection," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, ser. ETRA '12. New York, NY, USA: ACM, 2012, pp. 305–308. [Online]. Available: <http://doi.acm.org/10.1145/2168556.2168622>
- [55] Q. He, X. Hong, X. Chai, J. Holappa, G. Zhao, X. Chen, and M. Pietikäinen, "Omeg: Oulu multi-pose eye gaze dataset," in *Image Analysis*, R. R. Paulsen and K. S. Pedersen, Eds. Cham: Springer International Publishing, 2015, pp. 418–427.
- [56] S. Asteriadis, D. Soufleros, K. Karpouzis, and S. Kollias, "A natural head pose and eye gaze dataset," in *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots*, ser. AFFINE '09. New York, NY, USA: ACM, 2009, pp. 1:1–1:4. [Online]. Available: <http://doi.acm.org/10.1145/1655260.1655261>
- [57] S. D. Iyer and H. Ramasangu, "Hybrid lasso and neural network estimator for gaze estimation," in *2016 IEEE Region 10 Conference (TENCON)*, Nov 2016, pp. 2579–2582.
- [58] S. Nie, M. Zheng, and Q. Ji, "The deep regression bayesian network and its applications: Probabilistic deep learning for computer vision," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 101–111, Jan 2018.
- [59] E. Wood, T. Baltrušaitis, L.-P. Morency, P. Robinson, and A. Bulling, "Learning an appearance-based gaze estimator from one million synthesised images," in *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, ser. ETRA '16. New York, NY, USA: ACM, 2016, pp. 131–138. [Online]. Available: <http://doi.acm.org/10.1145/2857491.2857492>
- [60] E. Wood, T. Baltrušaitis, X. Zhang, Y. Sugano, P. Robinson, and A. Bulling, "Rendering of eyes for eye-shape registration and gaze estimation," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ser. ICCV '15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 3756–3764. [Online]. Available: <http://dx.doi.org/10.1109/ICCV.2015.428>
- [61] T. Baltrušaitis, P. Robinson, and L. P. Morency, "Openface: An open source facial behavior analysis toolkit," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2016, pp. 1–10.
- [62] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 2242–2251.
- [63] X. Zhang, Y. Sugano, and A. Bulling, "Everyday eye contact detection using unsupervised gaze target discovery," in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST 2017. New York, NY, USA: ACM, 2017, pp. 193–203. [Online]. Available: <http://doi.acm.org/10.1145/3126594.3126614>
- [64] S. J. Baek, K. A. Choi, C. Ma, Y. H. Kim, and S. J. Ko, "Eyeball model-based iris center localization for visible image-based eye-gaze tracking systems," *IEEE Transactions on Consumer Electronics*, vol. 59, no. 2, pp. 415–421, May 2013.
- [65] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [66] S. Bazrafkan, H. Javidnia, and P. Corcoran, "Face synthesis with landmark points from generative adversarial networks and inverse latent space mapping," *arXiv preprint arXiv:1802.00390*, 2018.
- [67] J. Lemley, S. Bazrafkan, and P. Corcoran, "Smart augmentation learning an optimal data augmentation strategy," *IEEE Access*, vol. 5, pp. 5858–5869, 2017. [Online]. Available: <https://doi.org/10.1109/ACCESS.2017.2696121>
- [68] —, "Learning data augmentation for consumer devices and services," in *Consumer Electronics (ICCE), 2018 IEEE International Conference on*. IEEE, 2018, pp. 1–3.
- [69] —, "Transfer learning of temporal information for driver action classification," in *The 28th Modern Artificial Intelligence and Cognitive Science Conference (MAICS)*, 2017.



**Joseph Lemley** Joseph Lemley received a B.S. degree in computer science and the Masters degree in computational science from Central Washington University in 2006 and 2016, respectively. He is currently pursuing a Ph.D. with the National University of Ireland Galway. His field of work is machine learning using deep neural networks for tasks related to computer vision. His Ph.D. is funded by FotoNation, Ltd., under the IRCSET Employment Ph.D. Program.



**Anurada Kar** Anuradha Kar is currently pursuing the Ph.D. degree with the National University of Ireland Galway and the Center for Cognitive, Connected, and Computational Imaging. Her research interests include human computer interaction and computational imaging. She is involved in eye gaze tracking-addressing the issues of accuracy and performance evaluation of gaze estimation systems in various platforms.



well as more than 30 patents.

**Alexandru Drimborean** Alexandru Drimborean is the Vice President of Advanced Research team at FotoNation Ireland focusing on developing innovative computer vision and machine learning technologies for mobile, biometrics and automotive applications. Alexandru received his B.S in Electronic Engineering in Brasov Romania followed by an M.Sc. in Electronic Science at N.U.I Galway in 2002. His interests include image processing and understanding as well computer vision and machine learning. Alexandru has authored several journal articles as



**Peter Corcoran** Peter Corcoran is a Fellow of IEEE, past Editor-in-Chief of IEEE Consumer Electronics Magazine and a Professor at NUI Galway. His research interests include biometrics, imaging, deep learning, edge-AI and consumer electronics. He is co-author on 350+ technical publications and co-inventor on more than 300 granted US patents. In addition to his academic career, he is an occasional entrepreneur, industry consultant and compulsive inventor.