TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAM	Active Appearance Model
AU	Action Unit
CAPD	Cornell Assessment of Pediatric Delirium
CAPP	Canonical-normalized Appearance
CERT	Computer Expression Recognition Toolbox
CHUSJ	Centre hospitalier universitaire Sainte-Justine
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
COMFORT-B	Comfort Behavioural
FACS	Facial Action Coding System
FLACC	Face Legs Activity Cry Consolability
GPU	Graphical Processing Unit
ICU	Intensive Care Unit
KNN	k-Nearest-Neighbours
LBP	Local Binary Pattern
LTP	Local Ternary Pattern
PCAM	Pediatric Confusion Assessment Method
PICU	Pediatric Intensive Care Unit
SAPP	Similarity-normalized Appearance

SPTS	Similarity-normalized Shape
SVM	Support Vector Machine

INTRODUCTION

Clinical context

Consciousness–that is, the patient's wakefulness and awareness of his or her surroundings–is an important indicator of health and neurological state. Altered consciousness, ranging from confusion and lethargy to loss of consciousness or even coma, may indicate many problems, including neurological disorders, poisoning, or brain injury (Avner, 2006). In cases when consciousness is purposefully altered by sedation or anesthesia, it is important to track the level of consciousness and ensure that patients are not oversedated or undersedated. Oversedation "will often result in prolonged mechanical ventilation and hemodynamic instability" (Johansson & Kokinsky, 2009), while undersedation may cause patients undue stress and pain.

It is therefore common practice to evaluate patient consciousness and distress regularly in intensive care units, using well-known tools like the Glasgow Coma Score (Teasdale & Jennett, 1974). In pediatric units, tools like the COMFORT behavioural (COMFORT-B) scale (Van Dijk, Peters, Van Deventer & Tibboel, 2005) and the Face Legs Activity Cry Consolability (FLACC) scale (Merkel, Voepel-Lewis, Shayevitz & Malviya, 1997) attempt to quantify consciousness and distress in children by observing their behaviour. These scales are typically evaluated by nursing staff using visual observation and require less than 5 minutes to perform, but the parameters of these scales are not always easy to score. For example, when determining the facial tension score for the COMFORT-B scale, it may be difficult for a nurse to distinguish between "facial muscle totally relaxed" and "facial muscle tone normal", especially for a child he or she has not seen before. As a result, different people may calculate different scores for the same patient.

The assessment of consciousness and distress in children under two years of age presents a particular challenge because these patients do not have the ability to respond to commands or explain what they are feeling. A study by Johansson & Kokinsky (2009) demonstrated that in

20% of cases there was a disagreement between nurses' bedside assessments of the level of sedation of young patients, highlighting the necessity of pediatric assessment tools to support their work.

The pediatrics research team at the Sainte-Justine University Hospital Centre (CHU Sainte-Justine) in Montreal is interested in developing a clinical decision support system that uses bedside cameras to detect signs of consciousness and distress due to pain.

Visual inspection of patients is an important part of clinical work and evaluation, yet the use of video for patient care remains rare in the hospital setting. Where video is used, it is often for retrospective analysis or remote monitoring by care staff rather than automated, real-time patient monitoring.

The development of compact, high-resolution cameras allows detailed and discreet monitoring directly at the point of care, and together with automated video image analysis would enable standardized, round-the-clock surveillance of patient state. This could be particularly valuable in situations where continuous bedside monitoring by trained staff is not possible, such as in remote regions or during medical transport. Such a system would also reduce human subjectivity in the evaluation of the COMFORT-B and FLACC scales and ensure the continuous and accurate recording and transmission of information throughout the patient's care.

The appearance and behaviour of the eye region are of particular interest in such a system because it can provide a wealth of information about a patient's condition and consciousness. Computer analysis of images using machine learning techniques has already allowed the identification of infant facial expressions and pain, as in Parodi, Melis, Boulard, Gavelli & Baccaglini (2017) where infant facial features like "eye squeeze" were detected in order to automatically score pain scales, and Fotiadou, Zinger, Ten, Oetomo & With (2014) where changes in facial appearance and expression were used to train a classifier to distinguish between pain and no-pain states.

The subjects in these studies were generally healthy with no feeding or breathing tubes or other facial occlusions, and they were recorded facing the camera.

Reliable localization and tracking of the eye region become more challenging when continuously monitoring directly in a hospital room, where patient age, condition, position and surroundings can vary widely. Patients may not be looking directly at the camera, and their faces may be partially occluded by medical equipment like nasal tubes or bandages. In addition, many existing approaches are developed and tested on adult facial datasets, which do not generalize well to infants and young children. Existing approaches and best practices for eye localization must therefore be re-examined for application in a pediatric clinical environment.

Research objectives

The goal of this work is to develop effective techniques for extracting the eye region from images of hospitalized pediatric patients, taking into account the challenges and constraints of the setting. Existing state-of-the-art computer vision and machine learning algorithms will be evaluated and refined for the task of detecting and tracking the eye region of children in a live clinical environment. What are the challenges and constraints of these techniques? How can they be refined to detect the eye region in the pediatric hospital setting? What datasets and technologies must be developed to support this project?

In parallel with technological development, a dataset is developed comprising videos of patients in the pediatric intensive care unit at CHU Sainte-Justine. These videos will be recorded with a photography camera pointed at the patient's upper body to acquire clear, high-resolution video footage of the facial region. Patients of all ages and medical conditions will be targeted for recording, so long as their eye region is at least partly visible. This dataset will fill an important gap for several research projects at CHU Sainte-Justine, as there are currently no publicly available datasets of hospitalized children available to researchers wishing to train and evaluate computer vision and machine learning models.

Achievements

The outcomes of this work demonstrate the potential of convolutional neural networks (CNNs) for eye localization in a pediatric hospital setting and the significant performance improvements that are gained by including a moderately sized sample of task-specific data during training. The convolutional neural network trained in this way performs better at eye localization than other methods on the novel dataset of 59 videos recorded of patients in the pediatric intensive care unit at Sainte-Justine Hospital. This real-world dataset demonstrates the challenges of working in a pediatric hospital environment, where faces are often partially occluded by positioning or medical devices.

CHAPTER 1

LITERATURE REVIEW

1.1 Object detection

Cascade classifiers, as first proposed in Viola & Jones (2001), are a widely used method for face and eye localization due to the availability of pre-trained models and the ease of training new models using limited data. Cascade classifiers are trained using manually defined features, such as Haar features or a local binary pattern (LBP), and use a sliding window approach to locate objects in new images.

The "cascade" refers to the narrowing of the search region as the classifier is applied to an image. Rather than calculate all relevant features for each window, a window is evaluated with a subset of features. If the object is not detected using those features, the window is discarded and not evaluated further. If, on the other hand, the results are promising, another subset of features is applied, and so on until the window is classified. This narrowing of the search region allows a classifier trained with a large number of features to run efficiently. Cascade classifiers are much faster than convolutional neural networks, making it easy to achieve real-time object localization.

The use of cascade classifiers is illustrated in a study to detect fatigue in bus drivers (Mandal, Li, Wang & Lin, 2017) which shares similar constraints to our project: images are recorded using a camera placed off-centre and above the subject, who may move freely and even face away from the camera. A chain of cascade detectors is used to identify first the upper body, then the face and its orientation, and finally the eye region. This gradual narrowing of the region of interest improves performance and accuracy by shrinking the candidate search region and eliminating false positives in the background.

Mingxin, Yingzi & Xiangzhou (2016) also use a trained classifier to search for eyes but add some rules: the search is limited to the upper half of the face, and this upper region is further divided into left and right halves (which should have one eye each). A variance filter is used to eliminate non-eye regions, then a support vector machine (SVM) is used to check the remaining candidate regions for eyes. El Kaddouhi, Saaidi & Abarkan (2017) eschewed the usual classifier and instead detected corner points using the Shi-Tomasi corner detector and grouped these points (by k-means) to produce candidate eye regions, which were then analyzed by template matching.

One approach not based on machine learning is proposed by Chen & Liu (2011). The facial image is converted to YUV colour space and the U component is thresholded to highlight the eye, which has a lower intensity than the surrounding region. Projection functions are then used on this binary image to find the eye boundaries. Chun-Ning, Tai-Ning, Pin & Sheng-Jiang (2012) describe another similar approach, without the thresholding, where rapid changes of intensity around the eyes are detected. However, both these approaches work best if the eyes are open, limiting their usefulness in practice.

Nevertheless, colour can be useful for narrowing the search region for other face and eye localization algorithms. For example, masking parts of an image based on skin colour can be used as a pre-processing step before a cascade classifier or other machine-learning-based approach, as in Mutneja & Singh (2017). It can also be used after face localization to discard face candidates that do not contain a sufficient number of skin colour pixels, as suggested by Ge, Han & Quan (2015). The accuracy of skin localization has been demonstrated on adults using the HSV and YCbCr colour spaces (Shaik, Ganesan, Kalist, Sathish & Jenitha, 2015), the RGB, YCbCr, and HSV colour spaces (Kolkur, Kalbande, Shimpi, Bapat & Jatakia, 2017), and the YCbCr colour space alone (Emmanuel & Ibiyemi, 2017; Ge *et al.*, 2015).

Convolutional neural networks (CNNs), a class of deep neural network particularly well suited to computer vision tasks, grew in popularity through the 2010s as computation power increased to permit faster training and predictions.

CNNs can be trained to sort images of objects into different classes, but there is an additional problem to solve in real-world situations: how to identify which objects are of interest in a crowded scene. Many algorithms have been proposed to solve this problem of object detection, among them Faster R-CNN (Ren, He, Girshick & Sun, 2015), which sacrifices the speed of other algorithms like YOLO (Redmon, Divvala, Girshick & Farhadi, 2016) for greater accuracy. Faster R-CNN builds on previous work in the R-CNN and Fast R-CNN algorithms, which used slow selective search algorithms to identify regions of interest in an image. Faster R-CNN speeds up this procedure by training a "region proposal network" during the training process, which identifies these regions of interest. Once identified, these subsections of a larger image can be classified by a conventional CNN.

CNNs underlie the most powerful facial and eye detection systems now available, and recent work has applied them in the hospital setting to detect adult patients exiting their beds (Chwyl, Chung, Shafiee, Fu & Wong, 2017), to identify the pose of adult patients in hospital beds (Liu, Yin & Ostadabbas, 2019a), and to detect infants in bed and segment their skin region (Chaichulee, Villarroel, Jorge, Arteta, Green, McCormick, Zisserman & Tarassenko, 2017). Nevertheless, the application of CNNs to real-world problems remains limited by the need for large quantities of training data and the computational cost of analyzing images through complex, many-layered networks.

1.2 Eye detection for pediatrics

Many of the approaches described above perform poorly when applied to young children in a hospital setting. As discussed by Saeijs, Tjona Ten & De With (2018), not only is there a lack of "in-the-wild" datasets of young hospitalized children with which to train models, but infant faces differ in appearance from those of older children and adults, complicating the use of models trained on adult datasets. Infant faces have less prominent features like eyebrows and skin folds, and facial proportions and morphology are different, making template and rules-based approaches more difficult to apply. Additionally, in a hospital setting faces may be partially occluded by items like bandages and tubes, and may move quickly and unpredictably.

Past studies interested in the facial features of babies and young children have largely avoided the problem of face and eye localization by manually selecting facial landmarks, as in Zamzami, Ruiz, Goldgof, Kasturi, Yu & Ashmeade (2015), or by using laborious alternative approaches like custom active appearance models, as in Fotiadou *et al.* (2014). Recent work in Chouinard, Scott & Cusack (2019) uses the Amazon Rekognition system to analyse infant facial images. Amazon Rekognition is a cloud-based computer vision platform that allows users to submit their own images for facial detection and analysis using pre-trained models. While results were promising, submitting patient data to an online service such as Amazon Rekognition raises privacy concerns that make this approach inappropriate for a medical setting.

1.3 The eyes as a measure of mental activity

Chen & Epps (2013) proposed several features that could be used to measure cognitive load in adults, such as pupil size, blink number, and saccade amplitude. They observed that the number of blinks and pupil size increased with more demanding tasks. Chen, Epps & Chen (2013) identified 29 possible eye features and used them to detect task transition (for which the most sensitive features were pupil response and blinks), perceptual load (for which the most sensitive features were blinks and saccades/fixations), and cognitive load (for which the most sensitive feature was average pupil size).

Several studies have tried to detect consciousness and alertness in car drivers. Kojima, Kozuka, Nakano & Yamamoto (2001) found that it is possible to detect alertness from the duration of blinks (longer blinks indicate drowsiness) and concentration from the variance of gaze movements (variance decreases during mental work). Jo, Lee, Park, Kim & Kim (2014) measured drowsiness based on percentage of frames with eyes closed in a given period. Tokuda, Obinata, Palmer & Chaparro (2011) showed that saccadic eye movements increase as mental work increases.

Blink rate decreases during visually demanding tasks (e.g., reading) but increases if a task is prolonged or made more difficult. (Stern, Boyer & Schroeder, 1994)

De Rivecourt, Kuperus, Post & Mulder (2008) showed that fixation duration and dwell time decreased as task demands increase during a flight simulation exercise.

1.4 Machine learning from eye movements

Lee, Ojha & Lee (2015a) fed four eye features into a one-class SVM classifier to determine concentration: number of eye blinks, duration of "open eye" state, pupil size variation, and rate of missing data (user looking away from eye tracker). The classifier was able to differentiate between concentration and non-concentration states for the small group of adult test users.

Jang, Mallipeddi, Lee, Kwak & Lee (2014) used fixation length, fixation count and pupil size variation to classify whether users were searching for an object or not when viewing images of indoor and outdoor scenes. They achieved a recognition rate of 85.26% using an SVM classifier.

1.5 Eye features

Table 1.1 presents a list of measurable features that can be detected in the eye region using computer vision techniques, given images or video of sufficient illumination and resolution. The exact resolution required varies depending on the feature: for example, a higher resolution is needed to detect changes in pupil size than to detect the opening and closing of the eye.

1.6 Consciousness in pediatrics

The pediatric age range presents a particular challenge for determining consciousness and alertness because infants and young children lack the ability to speak or follow instructions, their motor control is still poor, and it can be difficult to differentiate reflex responses from conscious responses.

Four major states of wakefulness are typically identified in newborns, each with unique eye behaviour. Table 1.2 lists these states and a brief description of the associated eye behaviour. (Yasova Barbeau & Weiss, 2017)

Various scales have been developed to quantify the neurological state of young children in a hospital setting, many using information from the face and eyes. Table 1.3 summarizes some of the most well-known coma, sedation, and delirium scales for infants and young children, along with a brief description of each and some examples of their criteria in and around the eye region.

These scales are reproduced in full in Appendix I: Selected pediatric evaluation scales.

Unlike the measurable eye features proposed in Table 1, these pediatric evaluation scales rely heavily on qualitative measures–for example, whether the eyes are "bright" or "dull", as in the Vancouver Sedative Recovery Scale. Evaluation is also limited by the abilities of infants and young children. For example, a newborn's verbal response to stimuli, which is a criterion measured by the Glasgow Coma Score, would be limited to cries.

Nelson, Lachman & Gold (2017) compared bedside diagnoses of delirium using the Cornell Assessment of Pediatric Delirium (CAPD) and Pediatric Confusion Assessment Method for the ICU (PCAM) to diagnoses performed by psychiatrists and found wide gaps between assessments by the two groups. This demonstrates the great difficulty of diagnosing cognitive state in pediatric patients, and the need for more accurate tools to assist clinicians.

Several studies have used eye tracking systems to gain insight into the cognitive development of young children (Corbetta, Guan & Williams, 2012; Saez de Urabain, Nuthmann, Johnson & Smith, 2017), but these studies are performed with alert children and eye trackers that require system calibration, something impossible to do with unconscious, distressed, or very young patients in a pediatric intensive care (PICU) environment.

1.7 Pediatric pain recognition

Parodi *et al.* (2017) proposed a tool that could automatically recognize infant facial features (for example, "eye squeeze" or "mouth stretch") that are used to calculate three traditional pain scores. The scores obtained from the system were compared to those calculated manually by users. There were high mismatch rates between the automatic and manual evaluation of scores (averaging near 50% mismatch for many parameters), but the study did not try to validate whether this was due to system error or user error.

Brahnam, Chuang, Sexton & Shih (2007a) photographed 26 neonates undergoing three unpleasant but non-painful stimuli (transport from one crib to another, a puff of air to the nose, friction on the outside of the heel) and one painful stimulus (a heel lance). The grayscale pixel intensity of the face region was extracted and the resulting feature vector was processed by four different machine learning techniques to classify the images as pain or nonpain. The highest classification accuracy achieved was 90.2%, but it was observed that certain subjects were harder to classify than others: all four algorithms performed poorly on the same subset of faces.

Nanni, Brahnam & Lumini (2010a) used the same dataset as the study above (Brahnam *et al.*, 2007a) to extract the following texture descriptors:

- **Local Binary Pattern (LBP)**: compares a pixel to each of its 8 neighbours and assigns 0 if its value is greater than the neighbour and 1 otherwise, generating an 8-digit binary number for each pixel.
- **Local Ternary Pattern (LTP)**: similar to LBP but assigns 0 if a pixel is the same value as its neighbour (within a threshold t), and -1 if it is greater than or -1 if it is less than the neighbour. This descriptor is less sensitive to noise than LBP.
- **ELBP/ELTP (Elongated LBP/LTP)**: uses an elliptical neighbourhood instead of circular. Not rotation invariant but can work better with images that have anisotropic structures (like faces).
- **ILBP/ILTP (Improved LBP/LTP)**: compares pixel neighbours to the local mean instead of the central pixel of a region. This descriptor is less sensitive to noise.

These texture descriptors were also extracted from images preprocessed using Gabor filters, a Laplacian of Gaussian filter, and the Illumination Normalization method, but the best classification results were obtained using grey levels. The best classification performance with an SVM (0.926 AUC) is obtained using the ELTP descriptor.

The above two approaches are limited by the static images used and cannot take into account facial movement and dynamic facial expressions. (Zamzni, Pai, Goldgof, Kasturi, Sun & Ashmeade,

2016) Furthermore, pixel intensity, as a feature, is sensitive to illumination changes and occlusion (for example, by an infant's hands). Texture descriptors like local binary patterns are less sensitive to illumination and noise (though still affected by occlusion) and may be well suited to detecting the wrinkles and furrows on the facial expressions of newborns. (Nanni *et al.*, 2010a)

Zamzami *et al.* (2015) collected video sequences for 9 neonates experiencing acute (heel lance) and chronic (post-operative recovery) pain. Facial landmarks were identified manually due to problems with existing detection algorithms, which are developed and trained on adult faces and do not deal well with unpredictable infant movements and occlusion by pacifiers and hands. The identified face is divided into four regions and an optical flow vector is generated for each of these regions and used to estimate optical strain. The estimated strain values for each region are added together to generate overall strain magnitude, which is used to classify the expression as pain or no-pain using k-Nearest Neighbours (KNN) and SVM classifiers. The KNN algorithm with k=3 produced the best classification accuracy of 96%.

Fotiadou *et al.* (2014) filmed video of 10 infants in a neonatal intensive care unit experiencing painful (heel lance) and non-painful (diaper change, hunger, resting, sleeping) experiences. During initialization, the infant's face is detected, its pose estimated (frontal or rotated), and it is fit to a pre-trained facial model (the Active Appearance Model, or AAM). This shape is then used to initialize future frames. The AAM allows the face's geometry to be manipulated to extract the following features:

- SPTS (similarity-normalized shape): the shape/geometry of the face in the current frame compared to the mean/default shape.
- SAPP (similarity-normalized appearance): the pixels of the facial image warped to the similarity-normalized shape.
- CAPP (canonical-normalized appearance): the pixels of the facial image warped to the mean shape.

Preprocessing techniques like Illumination Normalization and Laplacian of Gaussian filtering are used on the two appearance representations (SAPP and CAPP) before extracting features using various texture descriptors (including LBP and ELBP as described earlier), and an SVM classifier is trained to classify pain vs. no-pain. The best performance (0.98 AUC) is obtained using no preprocessing and ELBP texture descriptors on the appearance data. Performance with this approach is highly dependent on successful tracking of the face, which is challenging for infants whose movements are unpredictable and whose faces may be occluded by pacifiers or breathing tubes. Furthermore, custom AAMs were constructed for each infant in the study, but this is not a feasible approach for a real-world application given the time and effort required to create the models.

Sikka, Ahmed, Diaz, Goodwin, Craig, Bartlett & Huang (2015) recorded video of 50 post-operative youth (5-18 years old) experiencing pressure at a surgical site. Patients, parents, and nurses simultaneously rated pain using the Numerical Rating Scale. Videos were analyzed using the Computer Expression Recognition Toolbox (CERT) to extract standardized facial component movements known as facial action units (AUs), as described in the Facial Action Coding System (FACS). (Examples of AUs used in this study are "eye closure" and "upper lip raiser".) A 42-feature AU data vector was used to train two models:

- Binary classification (pain or no-pain), using a logistic regression model to learn mappings between features and binary pain labels.

- Pain-intensity estimation, using a linear regression model to learn a linear combination of features to predict pain self-ratings. A second pain-intensity model was trained using not only the AU vector but also "time since surgery" as a feature.

For binary classification, the trained model performed similarly to nurse and parent assessments of pain (>0.8 AUC). Performance was poorer for pain intensity estimation but comparable to the nursing assessment.

Table 1.1 Quantitative eye features

Feature	Quantitative measures
Blinks	• Number of blinks in a given time period • Duration of blinks • Interval between blinks
Fixations (gaze fixed on a single location)	• Number of fixations in a given time period • Duration of fixations • Interval between fixations • Direction of fixation
Saccades (quick gaze movements)	• Number of saccades in a given time period • Amplitude (how far does the gaze move each time?) • Variance (is the gaze moving around a lot or focusing on a small area?)
Pupils	• Pupil size • Pupil eccentricity • Accommodation reflex (changes in the eye when switching focus from a near to distant object, and vice-versa)
Eye region	• Eyebrows lowered, raised, or furrowed • Eyes open, closed, or squeezed shut
Response to stimulus	• Eye response to verbal stimuli • Eye response to pain • Delay before response

Table 1.2 Eye states by level of wakefulness

Name	**Eye state**
Quiet sleep	Eyes closed, no eye movements
Active sleep	Eyes closed, rapid eye movements
Transitional sleep/drowsy	Periods of opening and closing eyes, slow eye movements
Awake	Eyes open, rapid or slow eye movements

Table 1.3 Summary of pediatric evaluation scales

Name of scale	**Description**	**Examples of face- and eye-related criteria**
Modified Glasgow Coma Scale for Infants and Children (Teasdale & Jennett, 1974)	Widely used coma scale, adapted for infants and children	• Eye response to stimulus
COMFORT scale (Ambuel, Hamlett, Marx & Blumer, 1992)	Assesses sedation/distress in PICU patients	• Facial tension • Physical movements
Vancouver Sedative Recovery Scale (Macnab, Levine, Glick, Phillips, Susak & Elliott, 1994b)	Originally developed to assess recovery from major surgery	• "Bright" vs. "dull" eyes • Looks "at you" vs. "through you" • "Alert" vs. "flat" facial expression
Blantyre coma scale (Newton, Chokwe, Schellenberg, Winstanley, Forster, Peshu, Kirkham & Marsh, 1997)	Modified GCS developed to assess malarial coma in children	• Directed vs. not directed eye response
FLACC scale (Merkel *et al.*, 1997)	Pediatric pain assessment tool for children unable to communicate	• Facial expression
Richmond Agitation Sedation Scale (Sessler, Gosnell, Grap, Brophy, O'neal, Keane, Tesoro & Elswick, 2002)	Measures the agitation or sedation level of a patient	• Ability to maintain eyes open • Eye contact
Cornell Assessment of Pediatric Delirium (Traube, Silver, Kearney, Patel, Atkinson, Yoon, Halpert, Augenstein, Sickles & Li, 2014)	Assessment of pediatric delirium with developmental anchor points provided for scoring children under 2	• Awareness of surroundings • Eye contact
Preschool Confusion Assessment Method for the ICU (Smith, Gangopadhyay, Goben, Jacobowski, Chestnut, Savage, Rutherford, Denton, Thompson, Chandrasekhar et al., 2016)	Version of the Confusion Assessment Method adapted to young (nonverbal) children	• Response to pictures/mirrors • Ability to maintain eyes open

CHAPTER 2

ORGANIZATION OF THE DOCUMENT

The remainder of this document is structured as follows:

Chapter 3 describes the methodology of the work, providing background information and technical details about the algorithms and datasets that will be used in chapter 4.

Chapter 4 contains an article submitted for publication in the IEEE Journal of Biomedical and Health Informatics, covering the first phase of work done for CHU Sainte-Justine's broader project of developing a clinical decision support system that uses bedside cameras to detect signs of consciousness and distress due to pain. The article describes the development of a solution for eye localization, which involved assembling relevant training and test datasets, training cascade classifier models and convolutional neural networks for the task of eye localization in the pediatric hospital environment, and evaluating the results of these models against real-world data.

Chapter 5 discusses the implications of the results to the project at CHU Sainte-Justine.

Finally, the Conclusion summarizes the outcomes of this work and suggests avenues for future research and development.

Appendices provide supplementary material about the datasets and models created in this work: the new training dataset in Appendix II, the new test dataset in Appendix III, and the framework used to train convolutional neural networks in Appendix IV.

CHAPTER 3

METHODOLOGY

This project required the creation of new models for eye localization using machine learning methods. Before training these eye localization models, we also had to develop new datasets of training and test data since there were no suitable databases of images of babies and children available. We begin by introducing the datasets used or developed for this project, then the cascade classifiers and convolutional neural networks trained for the eye localization task, and finally the procedure and dataset used to evaluate the results.

3.1 Training Datasets

A total of five datasets were used for training the cascade classifiers and convolutional neural networks.

Four of these datasets are freely available on the internet for research purposes:

1. The Closed Eyes in the Wild dataset (Song, Tan, Liu & Chen, 2014)
   - 4848 images of open and closed eyes.
2. The annotated subset of the 10k US Adult Faces dataset (Bainbridge, Isola & Oliva, 2013)
   - 2222 annotated images of adult faces primarily with open eyes, cropped oval to remove a majority of hair and background.
3. The BioID Face Database (retrieved at https://www.bioid.com/facedb/)
   - 1521 annotated images of adult upper bodies in an interior setting, with varied facial expressions and lighting conditions.
4. Portions of the CIFAR-10 dataset (Krizhevsky & Hinton, 2009)
   - 4000 images of non-organic objects like ships and trucks.

We created a fifth dataset, consisting of 664 task-specific images of babies and young children in a medical setting, gathered using Google Images searches using the following search terms:

- Baby breathing tube,
- Baby eye hospital,
- Baby eyes,
- Baby hospital,
- Baby intubated,
- Baby NG tube.

This custom dataset adds a pool of setting- and task- specific images to the training dataset. No data recorded at CHU Sainte-Justine was used for training.

3.2 Algorithms

3.2.1 Cascade classifiers

Cascade classifiers have a long history of use as object detectors, especially for face and eye localization. Their efficiency and speed make cascade classifiers an obvious solution to explore for object detection problems.

To measure the effectiveness of cascade classifiers for our problem, we trained two cascade classifiers: one using LBP features and one using Haar features. All five datasets described in Section 3.1 were used, for a total 12,617 positive training images and 47,499 negative training images.

For the 10k US Adult Faces, BioID and custom datasets, positive training examples (images of eyes) were cropped from the full images using the eye annotations provided with the datasets, and negative training examples (images of non-eye regions and objects) were generated by randomly cropping sections of the face and background located outside the annotated eye regions.

All images were resized to 24x24 pixels, which corresponds to the smallest resolution of training images, those from the "Closed Eyes in the Wild" dataset. The images are in black and white, because training with colour features is not supported by the baseline OpenCV cascade classification algorithm.

The classifiers were trained using OpenCV's `opencv_traincascade` function until reaching precision >99.5%.

3.2.2 Convolutional neural networks

Convolutional neural networks, or CNNs, have grown in popularity in the past decade as hardware and architectural improvements have improved performance and reduced training time. However, as with all deep neural networks, large amounts of training data are needed to avoid overfitting. In our case, we have large amounts of training data of adult faces, but very little data related to children or to the hospital environment.

To measure the impact of different training data on performance and accuracy, we trained two convolutional neural networks.

The first CNN was trained using two datasets of adult faces, for a total of 3743 images:

1. The annotated subset of the 10k US Adult Faces dataset (Bainbridge *et al.*, 2013)
   - 2222 images of adult faces primarily with open eyes, cropped oval to remove a majority of hair and background.
2. The BioID Face Database (retrieved at https://www.bioid.com/facedb/)
   - 1521 images of adult upper bodies in an interior setting, with varied facial expressions and lighting conditions.

The second CNN was trained using the above two datasets, plus a third dataset of children's faces, for a total of 4407 images:

3. The custom dataset of images of babies and young children gathered from Google Images searches
   - 664 task-relevant images of babies and young children.

The two adult datasets were selected for the quality of their images and annotations. The high resolution of the full facial images ensures that the eye region will be sufficiently large and detailed for the CNN to extract good information during training. It also approximates the high resolution of the videos that make up our test dataset. The detailed annotations provided with these datasets identify the location of all major facial landmarks, including the eyes, allowing us to easily extract relevant regions for training.

Both CNNs were trained using Luminoth, a framework for training convolutional neural networks for object detection using the Faster R-CNN algorithm. (Ren *et al.*, 2015) This algorithm uses the training images and labels provided to produce two networks: a region proposal network that identifies objects of interest in a given image, and a classification network that classifies those objects of interest.

A ResNet-101 base network is used as a starting point for training the classification network. Deep residual networks, or ResNets, are a type of convolutional neural network whose architecture supports very deep (100+ layer) networks. (He, Zhang, Ren & Sun, 2016) ResNet-101 is a 101-layer network pre-trained on ImageNet (http://www.image-net.org), an image database currently containing links to more than 14 million images representing over 20,000 objects. The last two layers of this base network were fine-tuned during training with our provided images and annotations.

Images were used as-is during training, accompanied by the coordinates of the facial features in the image that the model will learn to locate. The thoroughness of the available facial landmark annotations led us to include both mouth and eye regions in our training. We thought that training the CNN to recognize and differentiate between these two facial features could improve eye localization performance, since it was observed when testing other eye localization solutions that many false positives occurred when mouths were mistaken for eyes. Mouths and eyes are typically the darkest regions on a light-coloured face. This, combined with the similar ellipsoid shape of the two facial features, can easily cause false detections, particularly in low lighting conditions or with low-quality images.

The training was allowed to run for 50 epochs, which was found in initial testing to minimize training loss without overfitting data or unduly extending training time.

3.2.3 Evaluation

Evaluation of the models produced was done using the dataset of babies and children recorded at CHU Sainte-Justine hospital. This dataset was not used during the training and validation of the developed models, making it ideal to verify the performance of the trained models and their ability to generalize to the real-world hospital setting.

Fifty-nine patients were recorded in their hospital rooms in CHU Sainte-Justine's pediatric intensive care unit between September 2018 and May 2019. A standard photography camera was used to capture videos with a resolution of 1920 x 1080 in RGB colour. In most cases, a single five-to-ten-minute video was recorded for each patient. Patients range from 9 days to 19 years of age, with approximately half (32) under the age of 2. Where possible, the camera was fixed at the lower-left or lower-right corner of the patient's bed, looking down at the upper half of the bed, and zoomed in to capture the full width of the bed.

Parental consent was obtained for all recordings and for the publication of the images included in this paper (CHU Sainte-Justine research ethical board approval number: 2016-1242).

Of the 59 patients recorded, one was discarded from the test set because the face was completely occluded by a breathing mask and one was discarded because the video was recorded while the patient was in a parent's lap resulting in multiple faces visible in the frame. Three patients under

6 months of age were filmed on different days or from different angles and contributed multiple recordings to the dataset.

Five still frames were randomly selected from each remaining video, for a total of 300 frames used for evaluation. The randomly selected frames were reviewed manually and those frames where the eye region was out of frame or completely occluded (for example, by the patient's hands) were discarded and replaced with another randomly selected frame. Frames with poor lighting, blurring, or other picture quality issues were kept, as long as the eye region was in-frame and visible.

Before evaluating images with the cascade classifiers, a skin colour filter was applied to mask areas of the image not likely to be the face or body. No such filter was applied to the images evaluated by the CNNs.

A test was considered successful if at least one eye was detected. This threshold was selected because both eyes were not visible in all images in the dataset, either due to patient positioning or occlusion by medical equipment. Given the dataset's small size, subdividing the data based on patient position and facial occlusion was not feasible, and automated identification of patient head position was outside the scope of this project.

CHAPTER 4

AUTOMATIC EYE LOCALIZATION FOR HOSPITALIZED INFANTS AND CHILDREN USING CONVOLUTIONAL NEURAL NETWORKS

Vanessa Prinsen[1,2] , Rita Noumeir[1] , Philippe Jouvet[2] , Sally Al Omar[2] , Gabriel Masson[2] , and Armelle Bridier[2]

[1] Département de génie éléctrique, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3
[2] Groupe de recherche clinique en soins intensifs pédiatriques, CHU Sainte-Justine,
3175 Chemin de la Côte-Sainte-Catherine, Montréal, Québec, Canada H3T 1C5

Paper submitted to the IEEE Journal of Biomedical and Health Informatics in January 2020.

Abstract

Reliable localization and tracking of the eye region in the pediatric hospital environment is a significant challenge for clinical decision support and patient monitoring applications. Existing work in eye localization achieves high performance on adult datasets but performs poorly in the busy pediatric hospital environment, where face appearance varies because of age, position and the presence of medical equipment.

A new training dataset, developed by gathering images of young children from internet searches, is used to train new cascade classifiers and convolutional neural networks for eye localization. Another new dataset, consisting of 59 recordings of patients in a pediatric intensive care unit, is used to evaluate the performance of these models. The convolutional neural network trained with the added image data of young children achieves a 79.7% eye localization rate, much higher than models trained on adult data alone. This model also outperforms the cascade models.

The effectiveness of convolutional neural networks, given the challenges of this setting, make it our preferred approach for eye localization and tracking in the hospital environment. The dramatic performance improvement gained from adding task-specific images to the training data highlights the need for custom-trained models for specialized applications like pediatric patient

monitoring. The moderate size of this added dataset is promising for future work, suggesting that it is feasible to develop an internal training dataset for clinical computer vision applications.

4.1 Introduction

Consciousness–that is, the patient's wakefulness and awareness of his or her surroundings– is an important indicator of health and neurological state. Altered consciousness, ranging from confusion and lethargy to loss of consciousness or even coma, may indicate many problems, including neurological disorders, poisoning, or brain injury (Avner, 2006). In cases when consciousness is purposefully altered by sedation or anesthesia, it is important to track the level of consciousness and ensure that patients are not oversedated or undersedated. Oversedation "will often result in prolonged mechanical ventilation and hemodynamic instability" (Johansson & Kokinsky, 2009), while undersedation may cause patients undue stress and pain.

It is therefore common practice to evaluate patient consciousness and distress regularly in intensive care units, using well-known tools like the Glasgow Coma Score (Teasdale & Jennett, 1974). In pediatric units, tools like the COMFORT-B scale (Van Dijk *et al.*, 2005) and the FLACC scale (Merkel *et al.*, 1997) attempt to quantify consciousness and distress in children by observing their behaviour. These scales are typically evaluated by nursing staff using visual observation and require less than 5 minutes to perform, but the parameters of these scales are not always easy to score. For example, when determining the facial tension score for the COMFORT-B scale, it may be difficult for a nurse to distinguish between "facial muscle totally relaxed" and "facial muscle tone normal", especially for a child he or she has not seen before. As a result, different people may calculate different scores for the same patient.

The assessment of consciousness and distress in children under two years of age presents a particular challenge because these patients do not have the ability to respond to commands or explain what they are feeling. A study by Johansson & Kokinsky (2009) demonstrated that in 20% of cases there was a disagreement between nurses' bedside assessments of the level of sedation of young patients, highlighting the necessity of pediatric assessment tools to support their work.

The pediatrics research team at the Sainte-Justine University Hospital Centre (CHU Sainte-Justine) in Montreal is interested in developing a clinical decision support system that uses bedside cameras to detect signs of consciousness and distress due to pain.

Visual inspection of patients is an important part of clinical work and evaluation, yet the use of video for patient care remains rare in the hospital setting. Where video is used, it is often for retrospective analysis or remote monitoring by care staff rather than automated, real-time patient monitoring.

The development of compact, high-resolution cameras allows detailed and discreet monitoring directly at the point of care, and together with automated video image analysis would enable standardized, round-the-clock surveillance of patient state. This could be particularly valuable in situations where continuous bedside monitoring by trained staff is not possible, such as in remote regions or during medical transport. Such a system would also reduce human bias in the evaluation of the COMFORT-B and FLACC scales and ensure the continuous and precise recording and transmission of information throughout the patient's care.

The appearance and behaviour of the eye region are of particular interest in such a system because it can provide a wealth of information about a patient's condition and consciousness. Computer analysis of images using machine learning techniques has already allowed the identification of infant facial expressions and pain, as in Parodi *et al.* (2017), where infant facial features like "eye squeeze" were detected in order to automatically score pain scales, and Fotiadou *et al.* (2014), where changes in facial appearance and expression were used to train a classifier to distinguish between pain and no-pain states. The subjects in these studies were generally healthy with no feeding or breathing tubes or other facial occlusions, and they were recorded facing the camera.

Reliable localization and tracking of the eye region become more challenging when we are continuously monitoring directly in a hospital room, where patient age, condition, position and surroundings may vary widely. The effectiveness of existing eye localization and tracking solutions must be re-examined for application in this environment.

In this paper, we present work done using a novel dataset of 59 videos recorded of patients in the pediatric intensive care unit at Sainte-Justine Hospital. Using this dataset, we demonstrate the potential of convolutional neural networks (CNNs) for eye localization in a pediatric hospital setting and the significant performance improvements that are gained by including a moderately sized sample of task-specific data during training. We also compare the performance of CNNs with that of cascade classifiers.

Our approach yields better results than other methods on our real-world dataset of pediatric patients, where faces are often partially occluded by positioning or medical devices. We discuss this outcome and the challenges and possibilities for eye localization techniques in a live clinical setting.

4.2 Technical challenges

Existing work in eye localization has achieved high performance on datasets of adults in controlled settings who exhibit a narrow range of facial variations such as glasses and facial hair.

In a live pediatric hospital setting, we must contend with greater differences in:

- **Age**: patients range from newborn to adult, which affects body and facial proportions. For example, the eyes are located lower in the face of an infant. These differences make facial landmark identification more challenging.
- **Position**: patients may be lying down, sitting up, facing the camera, or turned on their side.
- **Appearance**: in addition to typical variations in appearance, such as glasses, faces may also look different due to a medical condition or medical equipment, and facial features may be partially occluded by bandages, feeding tubes, or caregivers around the bed.

- **Environment**: although hospital rooms provided a more consistent environment than the real world, the area surrounding a patient can be visually complex, populated by medical equipment, toys, caregivers and visitors.

Additionally, any monitoring system must not impede clinical work. Mounting a camera in a suitable location, such as the ceiling or the foot of the bed, necessarily limits the resolution and quality of facial images that can be obtained.

As a result, we found very poor results when applying existing face and eye localization solutions, which are trained and tested primarily on controlled adult facial datasets, to real-world images of hospitalized pediatric patients. As shown in Table 4.1, the performance of existing face detection tools is significantly worse on a set of images of hospitalized children drawn from the 59 recordings made at Sainte-Justine Hospital: OpenCV's Haar cascade (Bradski & Kaehler, 2000) detected 83% of adult faces but only 33% of children's, while OpenFace (Baltrusaitis, Zadeh, Lim & Morency, 2018) correctly detected 100% of adult faces but only 45% of children's faces. The 42 test images of children were taken from the 59 video recordings made at CHU Sainte-Justine's pediatric intensive care unit over the course of this project. The 42 test images of adults come from the 10k US Adult Faces dataset.

Table 4.1 Performance of existing open-source face localization tools on 42 test images

Method	**Success rate**	
	Adults	Children
OpenCV built-in Haar cascade classifier	83% (35)	33% (14)
OpenFace	100% (42)	45% (19)

The development of a custom model is necessary, but the lack of publicly available datasets with annotated facial images of children and babies makes it difficult to train and test a new model. Acquiring such a dataset at the hospital is a slow process, limited by the unit's admission rate and parental consent. Additionally, while common wisdom is that large quantities of data are needed to train "deep" models such as CNNs, there are no precise guidelines on the quantity of data needed for good results. This makes it difficult to plan the development of a suitable training dataset.

4.3 Existing Work

Cascade classifiers, as first proposed in Viola & Jones (2001), are a widely used method for face and eye localization due to the availability of pre-trained models and the ease of training new models using limited data. Cascade classifiers are trained using manually defined features, such as Haar features or a local binary pattern (LBP), and use a sliding window approach to locate objects in new images.

The "cascade" refers to the narrowing of the search region as the classifier is applied to an image. Rather than calculate all relevant features for each window, a window is evaluated with a subset of features. If the object is not detected using those features, the window is discarded and not evaluated further. If, on the other hand, the results are promising, another subset of features is applied, and so on until the window is classified. This narrowing of the search region allows a classifier trained with a large number of features to run efficiently. Cascade classifiers are much faster than convolutional neural networks, making it easy to achieve real-time object localization, but their accuracy is limited because they are unable to learn new features and must use a predetermined set of features such as Haar edge features.

The use of cascade classifiers is illustrated in a study to detect fatigue in bus drivers (Mandal *et al.*, 2017) which shares similar constraints to our project: images are recorded using a camera placed off-centre and above the subject, who may move freely and even face away from the camera. A chain of cascade detectors was used to identify first the upper body, then the face and its orientation, and finally the eye region. This gradual narrowing of the region of interest improves performance and accuracy by shrinking the candidate search region and eliminating false positives in the background.

Convolutional neural networks (CNNs), a class of deep neural network particularly well suited to computer vision tasks, grew in popularity through the 2010s as computation power increased to permit faster training and predictions. Their application in real-world settings remains limited by the need for large quantities of training data and the computational cost of analyzing images through complex, many-layered networks. Nevertheless, CNNs underlie the most powerful facial and eye detection systems available today, and recent work has applied them in the hospital setting to detect adult patients exiting their beds (Chwyl *et al.*, 2017), identify the pose of adult patients in hospital beds (Liu *et al.*, 2019a), and detect infants in bed and segment their skin region (Chaichulee *et al.*, 2017).

Studies interested in the facial features of babies and young children have largely avoided the problem of face and eye localization by manually selecting facial landmarks, as in Zamzami *et al.* (2015), or by using time- and effort-intensive alternative approaches like custom active appearance models, as in Fotiadou *et al.* (2014). Recent work in Chouinard *et al.* (2019) uses the Amazon Rekognition system to analyse infant facial images. Amazon Rekognition is a cloud-based computer vision platform that allows users to submit their own images for facial detection and analysis using pre-trained models. While results were promising, submitting patient data to an online service such as Amazon Rekognition raises privacy concerns that make this approach inappropriate for a medical setting.

4.4 Methodology

This project required the creation of new models for eye localization using machine learning methods. Before training these eye localization models, we also had to develop new datasets of training and test data since there were no suitable databases of images of babies and children available. We begin by introducing two new datasets that we developed for this project, then describe the cascade classifiers and convolutional neural networks trained for the eye localization task.

4.4.1 Test data

Fifty-nine patients were recorded in their hospital rooms in CHU Sainte-Justine's pediatric intensive care unit between September 2018 and May 2019. A standard photography camera was used to capture videos with a resolution of 1920 x 1080 in RGB colour. In most cases, a single five-to-ten-minute video was recorded for each patient. Three patients under 6 months of age were filmed on different days or from different angles and contributed multiple recordings to the dataset. Parental consent was obtained for all recordings and for the publication of the images included in this paper (CHU Sainte-Justine research ethical board approval number: 2016- 1242). Patients range from 9 days to 19 years of age, with approximately half (32) under the age of 2. Where possible, the camera was fixed at the lower-left or lower-right corner of the patient's bed, looking down at the upper half of the bed, and zoomed in to capture the full width of the bed.

An example sequence of frames from one such video is shown in Figure 4.1.

Figure 4.1 Series of frames from one recording at CHU Sainte-Justine

Of the 59 patients recorded, one was discarded from the test set because the face was completely occluded by a breathing mask and one was discarded because the video was recorded while the patient was in a parent's lap resulting in two faces in the frame.

A set of still images was randomly extracted from the recordings of the remaining fifty-seven patients to form a test set for eye localization, hereafter referred to as the CHUSJ test set. Five frames were randomly selected from each recording, for a total of 300 still frames (including three patients with more than one recording available).

The randomly selected frames were reviewed manually and those frames where the eye region was out of frame or completely occluded (for example, by the patient's hands) were discarded and replaced with another randomly selected frame. Frames with poor lighting, blurring, or other picture quality issues were kept, as long as the eye region was in-frame and visible.

4.4.2 Training data

We also created a new dataset to help train our eye localization models, consisting of 664 relevant images of babies and young children gathered using Google Images searches using the following search terms:

- Baby breathing tube,
- Baby eye hospital,
- Baby eyes,
- Baby hospital,
- Baby intubated,
- Baby NG tube.

The eye and mouth regions in these dataset images were manually labelled using labelImg, an image annotation tool downloaded from Github (Git, 2015). The resulting images and annotations were included in our training data.

4.4.3 Cascade classifiers

Cascade classifiers have a long history of use as object detectors, especially for face and eye localization. Their efficiency and speed make cascade classifiers an obvious solution to explore for object detection problems.

To measure the effectiveness of cascade classifiers on our test data of children in a hospital setting, we trained a pair of cascade classifiers: one using LBP features and one using Haar features. Both cascade classifiers were trained using OpenCV's opencv_traincascade application with default parameters.

Five datasets were used for training:

1. The Closed Eyes in the Wild dataset (Song *et al.*, 2014)
   - 4848 images of open and closed eyes.
2. The annotated subset of the 10k US Adult Faces dataset (Bainbridge *et al.*, 2013)
   - 2222 images of adult faces primarily with open eyes, cropped oval to remove a majority of hair and background.
3. The BioID Face Database (retrieved at https://www.bioid.com/facedb/)
   - 1521 images of adult upper bodies in an interior setting, with varied facial expressions and lighting conditions.
4. Portions of the CIFAR-10 dataset (Krizhevsky & Hinton, 2009)
   - 4000 images of non-human objects like ships and trucks.
5. The custom dataset of images of babies and young children gathered from Google Images searches
   - 664 task-relevant images of babies and young children.

The complete training dataset for the cascade classifiers consisted of 12,617 positive (eye) and 47,499 negative (non-eye) training images, resized to 24x24 pixels. This corresponds to the resolution of the smallest training images used, those from the "Closed Eyes in the Wild" dataset. Images are in black and white, because training with colour features is not supported by the baseline OpenCV cascade classification algorithm.

For the 10k US Adult Faces, BioID and custom datasets, positive training examples (images of eyes) were cropped from the full images using the landmark annotations included with the dataset, and negative training examples (images of non-eye regions and objects) were generated by randomly selecting 24 x 24 pixel patches from the face and background outside the eye region.

Additionally, hard negative mining was used on the custom baby dataset and the 10k US Adult Faces dataset. Figure 4.2 illustrates the flow of training with hard negative mining. We train a cascade classifier using an initial set of images with positive (eye) and negative (non-eye) regions identified. Then, the trained cascade classifier is applied back to the training images to generate a set of eye region predictions. These are divided into correct predictions – predicted eye regions containing an eye – and incorrect predictions – predicted eye regions not containing an eye. The incorrect predictions are added back into the negative training data and a second version of the cascade classifier is trained. This approach allows us to fine-tune the classifier's performance using the data it finds most difficult to classify.

Before applying the cascade classifiers, we also pre-process our images to mask regions not matching a simple skin colour filter. This is done because the cascade classifiers are trained only to recognize object shape and do not take into account colour features, and so may have difficulty distinguishing between actual eyes and things that look like eyes like gauges on medical equipment and eyes on toys. A binary mask is generated by checking each image pixel against criteria in the RGB and YCbCr colour spaces and then applying morphological operations to fill holes and remove noise from the resulting mask. The following criteria are used, drawn from a combination of thresholding approaches summarized in the survey by Eastwood-Sutherland, Gale, Dargavillc & Wheeler (2016) :

- $R > G$

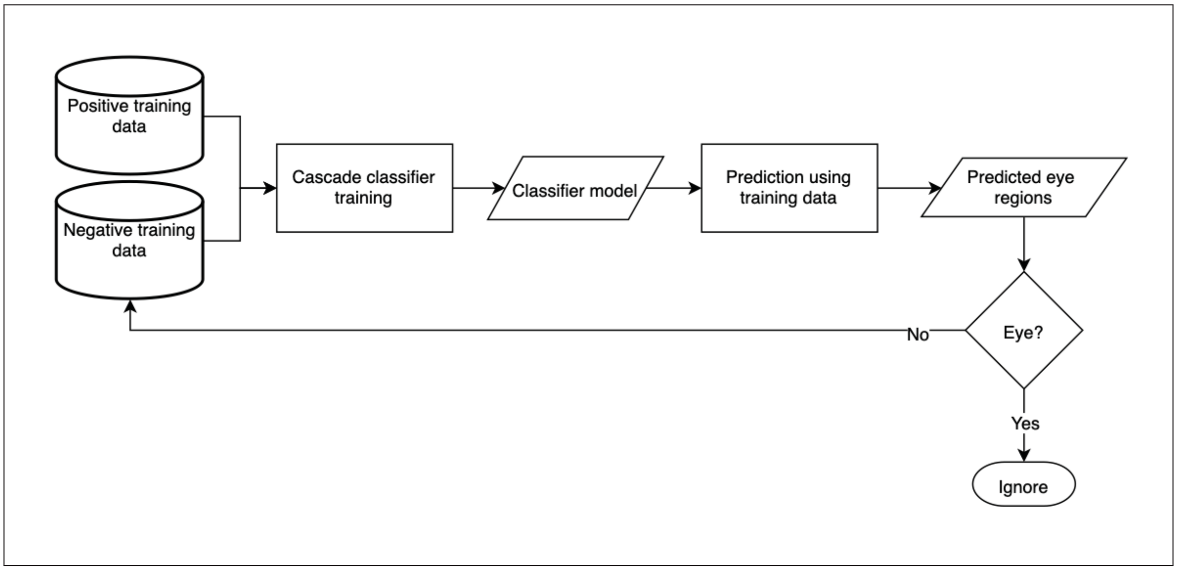

Figure 4.2 Hard negative mining

- R > B
- 70 < Cb < 135
- 130 < Cr < 170

4.4.4 Convolutional neural networks

Convolutional neural networks, or CNNs, have grown in popularity in the past decade as hardware and architectural improvements have improved performance and reduced training time. An architecture combining a region-proposal algorithm with convolutional neural networks, known as R-CNN, has spawned many derivatives that now constitute state-of-the-art approaches for object detection and segmentation in complex scenes. However, as with all deep neural networks, large amounts of training data are needed to avoid overfitting.

In our case, we have large amounts of training data of adult faces, but very little data related to children or to the hospital environment. We are therefore interested in measuring the impact of different training data on the performance of the convolutional neural network. To that end, we trained a pair of convolutional neural networks: one using only adult datasets, and one using adult datasets as well as our custom task-specific dataset drawn from Google Images searches of babies and young children.

Both convolutional neural networks were trained using the Luminoth framework (Tryolabs, 2018), which provides an implementation of the Faster R-CNN algorithm with ResNet-101 as a base network.

Deep residual networks, or ResNets, are a type of convolutional neural network whose architecture supports very deep (100+ layer) networks. (He *et al.*, 2016) The architecture achieved <5% error rates on the ImageNet test set and won several image detection and localization challenges in 2015. ResNet-101 is a 101-layer network trained on ImageNet (http://www.image-net.org), an image database currently containing links to more than 14 million images representing over 20,000 objects.

In our case, we use the ResNet-101 architecture and retrain all the network weights to fit our own training data. This training data consists of full images, each containing at least one face, accompanied by the coordinates of the facial features in the image that we want the model to learn to locate.

The training was allowed to run for 50 epochs, which was found in initial testing to minimize training loss without overfitting data or unduly extending training time. One model was allowed to train for 200 epochs, but this yielded no performance gain with our data and the 50-epoch model was used for final testing.

We trained our first convolutional neural network using two datasets of adult faces only:

1. The annotated subset of the 10k US Adult Faces dataset (Bainbridge *et al.*, 2013)
    - 2222 images of adult faces primarily with open eyes, cropped oval to remove a majority of hair and background.
2. The BioID Face Database (retrieved at https://www.bioid.com/facedb/)

- 1521 images of adult upper bodies in an interior setting, with varied facial expressions and lighting conditions.

We trained our second convolutional neural network using the same two adult datasets, plus a third dataset of children's faces:

3. The custom dataset of images of babies and young children gathered from Google Images searches
    - 664 task-relevant images of babies and young children.

We selected the two adult datasets for the quality of their images and annotations. The high resolution of the full images ensures that the eye region will be sufficiently large and detailed for the CNN to extract good information during training. It also approximates the high resolution of the videos that make up our test dataset. The annotations provided with the dataset identify the location of all major facial landmarks, including the eyes, allowing us to easily extract relevant regions for training.

The thoroughness of these annotations led us to include both mouth and eye regions in our training. We thought that training the CNN to recognize and differentiate between these two facial features could improve eye localization performance, since it was observed when testing other eye localization models that many false positives occurred when mouths were mistaken for eyes. Mouths and eyes are typically the darkest regions on a light-coloured face. This, combined with the similar ellipsoid shape of the two facial features, can easily cause false detections, particularly in low lighting conditions or with low-quality images.

4.5 Results

We use the CHUSJ test set, 300 still images of 57 patients from Sainte-Justine Hospital, to evaluate the performance of five eye localization models:

- The Haar cascade classifier included with OpenCV;

- Four custom-trained models:
  - Cascade classifier with LBP features,
  - Cascade classifier with Haar features,
  - Convolutional neural network with adult data only,
  - Convolutional neural network with adult and child data.

The resulting eye candidates are evaluated manually to determine the number of frames in which one of the patient's eyes was selected as the top localization result. We have chosen to consider only one eye in this evaluation because our dataset contains several recordings where the second eye is partially or fully occluded by medical condition or patient position.

Figure 4.3 illustrates the results of our evaluation of the five eye localization models.

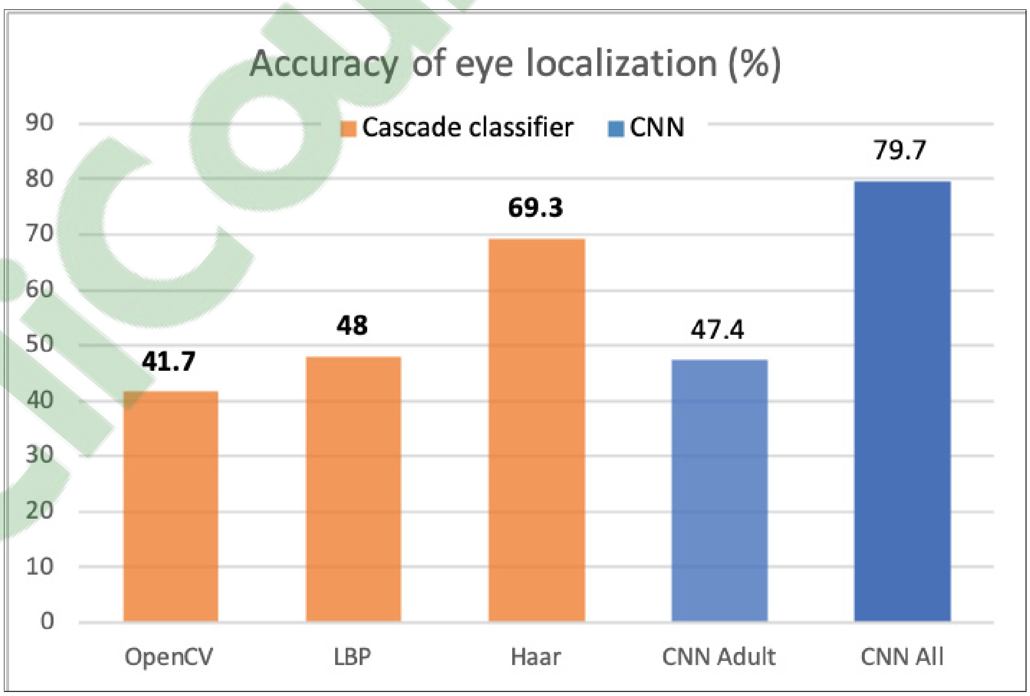

Figure 4.3 Accuracy of eye localization with cascade classifiers and CNNs

The best result for a cascade classifier was the model using Haar features which successfully located an eye in 69.3% of frames.

The convolutional neural network trained with only adult images performed as poorly as most of the cascade classifiers, with a localization success rate of only 47.4%. When the custom dataset of images of babies and young children gathered from Google Images searches was added to the training data, the accuracy of eye localization increased to 79.7%.

An additional performance gain was achieved by adding Contrast Limited Adaptive Histogram Equalization (CLAHE) as a pre-processing step. This yielded a promising 84% success rate. CLAHE divides an image into smaller tiles that are individually histogram equalized to prevent information loss due to large contrast variations across an image.

Figure 4.4 compares the top-2 localization results on sample frames from three different patients in the CHUSJ test set. The first column shows frames analyzed by the Haar cascade classifier with skin masking applied, the second column shows frames analyzed by the convolutional neural network trained only on adult datasets, and the third column shows frames analyzed by the convolutional neural network trained with adult datasets plus the custom dataset of images of babies and young children from Google Images searches. The region selected in green indicates the highest probability eye candidate, and the region selected in yellow indicates the second-highest probability eye candidate.

In the first two rows of images of babies, we see that the Haar cascade classifier has produced false positives. Overall, 75 out of 300 images identified an incorrect region of the image as the top eye detection result, for a false positive rate of 25%. The convolutional neural network is more restrained in locating eyes on these images. When trained with adult data only, it locates only the most visible, well-defined eye in the image of the first patient, and does not detect either of the closed eyes in the image of the second patient. In contrast, when trained with both adult and baby data, it locates all eyes in both images.

Performance, even with the Haar cascade classifier, is much better in the image of the third patient, who is older and in a relatively uncluttered environment. This illustrates the challenge of applying models trained and tested on adult datasets to young children and babies, particularly in a complex environment like a hospital room.

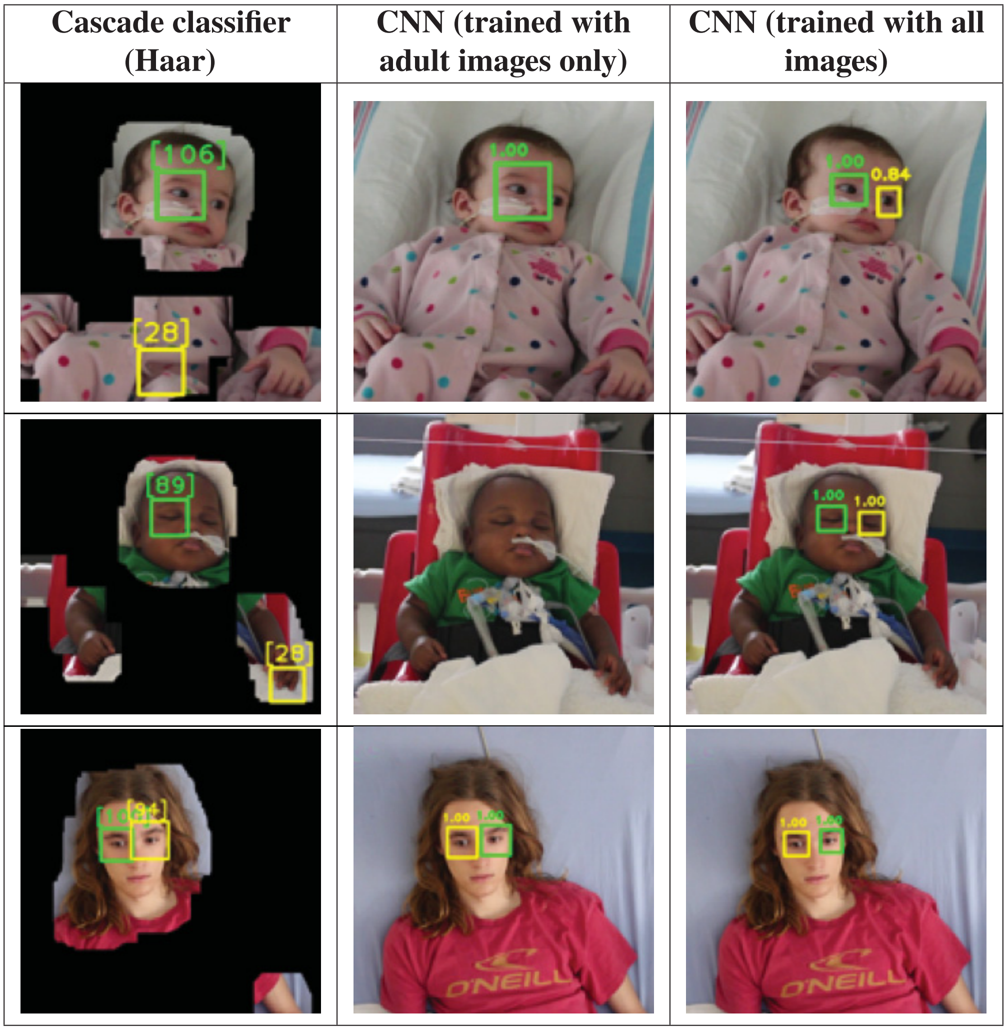

Figure 4.4 Comparison of eye localization results

4.6 Discussion

4.6.1 Task-specific training dataset

The dramatic performance improvement seen when we added the custom dataset of images of babies and young children to the training data of the convolutional neural network demonstrates the impact that a relatively small dataset of curated, task-specific images can have on training a model.

4.6.2 Model weaknesses

Closed eyes were generally harder to locate than open ones, as they can easily resemble folds and shadows elsewhere on the face and body. The localization failures noted with the CNN consisted of sleeping children with poorly defined or partially occluded eyes, children with their heads strongly tilted so that open eyes appeared closed and closed eyes appeared more vertical than horizontal, and heavily blurred frames. Some of these issues can be resolved by expanding the training dataset to include more images of sleeping children and by adding rotated images to the training set to improve recognition of eyes in various positions. In a real-time application, missing eye localizations could also be infilled based on previous frames, patching over momentary tracking drops due to blurring or other recording artefacts.

4.6.3 Convolutional neural networks vs. cascade classifiers

The significant difference in performance between convolutional neural networks and cascade classifiers can be explained by the features used in these two approaches. A convolutional neural network determines relevant features for object localization automatically during training, while a cascade classifier relies on human-defined features (such as Haar or LBP features).

Haar features, while highly effective for many applications, use a limited set of 90- and 45-degree edges for object detection. LBP features similarly look at local variations between orthogonal and diagonal neighbouring pixels. This method of object detection works well for recognizing landmarks on aligned, front-facing faces, as demonstrated in Viola & Jones (2001), where the eyes and mouth form horizontal edges, the nose forms a vertical edge, and so on. These features perform poorly on faces that are deformed, occluded, tilted or rotated, like those of children lying in a hospital bed. The classifier is also likely to be confused by the background clutter common in hospital rooms.

Though our results seem strongly in favour of convolutional neural networks, it should be noted that cascade classifiers currently perform much faster than CNNs. The slowest cascade classifier, with the added pre-processing step of skin masking, ran at a frame rate of approximately 5 fps on a laptop with no GPU acceleration. The same laptop took approximately 40s to analyze just one frame using the trained CNNs. A lighter-weight MobileNet network, developed specifically for mobile phone applications, has been used to run image classification at 5 fps (Garcia-Garcia, Caplier & Rombaut, 2018), so it is possible that a lightweight neural network architecture and a more powerful computer equipped with a GPU could achieve near-real-time performance on eye localization with a convolutional neural network. Ongoing improvements to both computer hardware and convolutional neural network architectures also make it likely that CNNs will be a viable solution for real-time object localization and image analysis tasks in the near future.

Finally, the convolutional neural networks demonstrated fewer false positives than the cascade classifiers. An incorrect eye detection occurred in only 2 of 300 images evaluated, for a false positive rate under 1%. This is advantageous in a hospital environment where it is better to alert clinicians that eye tracking has failed than to return incorrect information about a patient.

4.6.4 Limitations

Different hyperparameters, algorithms, etc.

Performance:

An alternate algorithm, YOLO, sacrifices some of its accuracy for speed. It was not evaluated in this work as the primary interest was accurate detection of the eye region, but it should be evaluated in the future as its speed would make real-time analysis of patient images more feasible.

Evaluation on still images only: no possiblity for frame-filling or inference possible.

4.7 Conclusion

Despite the challenges of bringing video monitoring and eye localization into the pediatric hospital environment, our work has demonstrated the potential of convolutional neural networks for eye localization in a pediatric hospital setting and the significant performance improvements that can be gained by including a moderately sized sample of task-specific data during training.

Using a novel dataset of 59 videos recorded of patients in the pediatric intensive care unit at Sainte-Justine Hospital, we compared the performance of convolutional neural networks with that of cascade classifiers and analyzed the impact of different training data on CNN performance. Although trained with a smaller dataset, our best convolutional neural network achieved much better performance than any cascade classifier we tested. Image preprocessing to correct contrast may be helpful to further improve performance.

The dramatic performance improvement gained from adding task-specific images to the training data highlights the need for custom-trained models for specialized applications like pediatric patient monitoring. The moderate size of this added dataset suggests that it is feasible to develop an internal training dataset for clinical computer vision applications.

Future work is needed to explore whether newer convolutional neural network architectures, such as the lightweight MobileNet, can achieve real-time performance. Further development of the training dataset, for example by including rotated and mirrored images, could also increase eye localization performance.

CHAPTER 5

DISCUSSION OF THE RESULTS

The objectives of this work were to develop effective techniques for extracting the eye region from images of hospitalized pediatric patients, taking into account the setting's challenges and constraints as described in section 4.2. A novel dataset of videos of children in a pediatric hospital environment was created and used to evaluate state-of-the-art computer vision and machine learning algorithms for the task of eye localization. Thanks to this dataset, it was found that existing tools and approaches, which perform extremely well on adult facial datasets, perform poorly on images of hospitalized infants and children. This finding shows the importance of developing internal datasets for specialized clinical applications and patient populations. The dataset developed as part of this work will be of ongoing importance to the clinical decision support project and other research in computer vision at CHU Sainte-Justine.

The dramatic performance improvement seen when adding a moderately sized dataset of images of babies and young children to the convolutional neural network's training data also sheds light on the feasibility of creating internal training datasets specifically for hospital applications. With several hundred of admissions each year, assembling a comparable 600-image training data set is within the reach of a large university hospital. When combined with existing public datasets of adult facial images, this task-specific training data produced a convolutional neural network with a 79.7% eye localization rate, far better than existing solutions trained only on adult data.

As pervasive camera-based monitoring becomes more common in hospitals, internal datasets of patient images will grow, enabling ongoing refinement of computer vision models. The current model's weaknesses at locating poorly defined eyes, closed eyes, and partially occluded eyes will diminish as the set of training data grows to encompass a wider range of eye appearances.

The above localization rate was also achieved with only one specific neural network architecture and using no pre-processing or information from previous frames. With further improvements, like various image preprocessing techniques and frame infilling, not to mention more advanced neural network architectures, it can be expected that model accuracy will soon reach levels suitable for real-world applications. As hardware performance and neural network architectures both improve, the processing speed of CNNs will also increase, permitting their use in real-time applications.

Overall, it can be seen that the convolutional neural network's ability to learn unique image features enabled it to adapt to the challenges of eye localization in a pediatric hospital environment, where the assumptions behind other approaches to face and eye localization do not necessarily apply. For example, the CNN naturally integrated colour information into its model, removing the need for the colour masking done before applying the cascade models. It also identified few false negatives, an important advantage in clinical applications where no results are better than false results about patient condition.

CONCLUSION AND RECOMMENDATIONS

There are many challenges to bringing video monitoring and eye localization into the pediatric hospital environment. Existing work in eye localization has achieved high performance on adult datasets but performs poorly in the busy pediatric hospital environment, where face appearance varies because of age, position and the presence of medical equipment. The goal of this work was to develop effective techniques for extracting the eye region from images of hospitalized pediatric patients, enabling future development of computer-vision-based clinical decision support systems.

A new training dataset, formed of images of young children from internet searches, was created and combined with existing datasets of adult facial images to train cascade classifiers and convolutional neural networks. Another novel dataset, consisting of 59 recordings of patients in a pediatric intensive care unit, was used to evaluate and compare the performance of these models.

This analysis demonstrated that convolutional neural networks can achieve superior performance at the eye localization task, even when trained with a smaller dataset. The impact of different datasets on model performance was also assessed, finding that significant performance improvements can be obtained by including a moderately sized sample of task-specific data during model training.

As has already proven the case in many other domains, convolutional neural networks have shown themselves to be powerful and promising tools for object detection in the pediatric hospital environment. Their performance at eye localization in the CHU Sainte-Justine dataset makes it possible to reliably identify the eye region in most recorded patient videos, enabling future work to calculate scores like the FLACC and COMFORT-B scales using computer vision and machine learning techniques.

Recommendations for future work are to:

- Continue developing the internal dataset of images and videos of hospitalized children. This data constitutes an important resource for training future models, and a vital contribution to other clinical research projects in computer vision given the present lack of publicly-available datasets of images of hospitalized children.
- Explore different neural network architectures, particularly lightweight ones like MobileNet, to improve object detection speeds and achieve real-time performance. Model performance should be tested on realistic hardware, such as what might be found in a hospital room or on a hospital server.
- Further develop the existing training dataset by including rotated and mirrored images, to improve performance in cases where patient position is unusual.
- Continue developing the eye localization and tracking application, to improve overall localization rates by infilling missing data from surrounding frames and equalizing image contrast.

APPENDIX I

SELECTED PEDIATRIC EVALUATION SCALES

Table-A I-1 Coma and sedation assessment tools for children

Name of scale	Evaluation criteria	
COMFORT-B scale (Van Dijk *et al.*, 2005)	**Alertness:** • Deeply asleep • Lightly asleep • Drowsy • Fully awake and alert • Hyperalert **Calmness/agitation:** • Calm • Slightly anxious • Anxious • Very anxious • Panicky **Physical movement:** • No movement • Occasional, slight movement • Frequent, slight movement • Vigorous movement limited to extremities • Vigorous movements including torso and head	**Facial tension:** • Facial muscles totally relaxed • Facial muscle tone normal, no facial muscle tension evident • Tension evident in some facial muscles • Tension evident throughout facial muscles • Facial muscles contorted and grimacing **Muscle tone:** • Muscle totally relaxed, no muscle tone • Reduced muscle tone • Normal muscle tone • Increased muscle tone and flexion of fingers and toes • Extreme muscle rigidity

Name of scale	Evaluation criteria	
Modified Glasgow Coma Scale for Infants and Children (Reilly, Simpson, Sprod & Thomas, 1988)	**Infants** **Eye opening:** • Opens spontaneously • Open in response to verbal stimuli • Open in response to pain only • No response **Verbal response:** • Coos and babbles • Irritable cries • Cries in response to pain • Moans in response to pain • No response **Motor response:** • Moves spontaneously and purposefully • Withdraws to touch • Withdraws in response to pain • Responds to pain with decorticate posturing (abnormal flexion) • Responds to pain with decerebrate posturing (abnormal extension) • No response	**Children** **Eye opening:** • Opens spontaneously • Open in response to verbal stimuli • Open in response to pain only • No response **Verbal response:** • Oriented, appropriate • Confused • Inappropriate words • Incomprehensible words or nonspecific sounds • No response **Motor response:** • Obeys commands • Localizes painful stimulus • Withdraws in response to pain • Responds to pain with decorticate posturing (abnormal flexion) • Responds to pain with decerebrate posturing (abnormal extension) • No response

Name of scale	Evaluation criteria	
Blantyre coma scale (Newton *et al.*, 1997)	**Eye response:** • 1: Directed eye movements • 0: Not directed **Verbal:** • 2: Appropriate cry • 1: Inappropriate cry/moan • 0: No cry	**Motor:** • 2: Localizes pain • 1: Withdraws from pain • 0: No response

Name of scale	Evaluation criteria
Face Legs Activity Cry Consolability (FLACC) scale (Merkel *et al.*, 1997)	**Face:** • 0: No particular expression or smile • 1: Occasional grimace or frown, withdrawn, uninterested • 2: Frequent to constant quivering chin, clenched jaw **Legs:** • 0: Normal position or relaxed • 1: Uneasy, restless, tense • 2: Kicking, or legs drawn up **Activity:** • 0: Lying quietly, normal position, moves easily • 1: Squirming, shifting, back and forth, tense • 2: Arched, rigid or jerking **Cry:** • 0: No cry (awake or asleep) • 1: Moans or whimpers; occasional complaint • 2: Crying steadily, screams or sobs, frequent complaints **Consolability:** • 0: Content, relaxed • 1: Reassured by occasional touching, hugging or being talked to, distractible • 2: Difficult to console or comfort

Name of scale	Evaluation criteria
Richmond Agitation Sedation Scale (Sessler *et al.*, 2002)	• +4 Combative, violent, immediate danger to staff • +3 Very agitated, pulls to remove tubes/catheters, aggressive • +2 Agitated, frequent non-purposeful movement, fights ventilator • +1 Restless, anxious, apprehensive, movements not aggressive • +0 Alert & calm, spontaneous awake • -1 Drowsy, not fully alert but sustained awakening to voice (eye opening and contact >10s) • -2 Light sedation, briefly awakens to voice (eye opening and contact <10s) • -3 Moderate sedation, movement or eye opening to voice (no eye contact) • -4 Deep sedation, no response to voice, movement or eye opening to stimulation • -5 Unarousable, no response to voice or physical stimulation

Name of scale	Evaluation criteria
Vancouver Sedative Recovery Scale (Macnab, Levine, Glick, Phillips, Susak & Elliott, 1994a)	**A: (Response)** • Awake/alert • Awake/drowsy • Asleep/easily aroused • Asleep/difficult to arouse • Asleep/unable to arouse **B: (Response)** • Responds fully to stimuli in an age-appropriate manner • Delayed response to stimuli • Absent response to stimuli **C: (Response)** • "Alert" facial expression • "Flat" facial expression **D: (Eyes)** • Bright eyes • Dull eyes; glazed **H: (Eyes)** • Purposeful and spontaneous eye movement • Little or no spontaneous or purposeful eye movement **I: (Movement)** • Spontaneous and varied central activity • Spontaneous and varied peripheral activity • Central activity w/ stimuli • Peripheral activity w/ stimuli • No movement **J: (Movement)** • Absence of tremor or ataxia • Minor ataxia or tremor • Major ataxia or tremor

Name of scale	Evaluation criteria	
Vancouver Sedative Recovery Scale (Macnab *et al.*, 1994a)	**E: (Eyes)** • Looks "at you" • Looks "through you" **F: (Eyes)** • Accommodates • Does not accommodate **G: (Eyes)** • Recognition of stimulus • Limited or no recognition	**K: (Movement)** • Coordinated spontaneous • Weak/coarse spontaneous • No purposeful sponta-neous **L: (Movement)** • Age-appropriate manual dexterity • Awkward or clumsy hand movement • No free hand movement

Table-A I-2 Delirium assessment tools for young children

Name of scale	**Evaluation criterias**
Cornell Assessment of Pediatric Delirium (Traube *et al.*, 2014)	• Does the child make eye contact with the caregiver? • Are the child's actions purposeful? • Is the child aware of his/her surroundings? • Does the child communicate needs and wants? • Is the child restless? • Is the child inconsolable? • Is the child underactive? Very little movement while awake? • Does it take the child a long time to respond to interactions? * To apply score to children under 2 years of age, developmental anchor points are provided for each criterion

Name of scale	**Evaluation criterias**
Preschool Confusion Assessment Method for the ICU (Smith *et al.*, 2016)	• Acute change or fluctuating course of mental status - Is there an acute change from mental status baseline? - Has the patient's mental health status fluctuated in the past 24 hours? • Inattention: show alternating pictures/mirrors and give verbal prompts - Attends to 7 or less pictures/mirrors? - Patient does not maintain spontaneous eye opening in between verbal prompts? • Altered level of consciousness (LOC) - Does the patient currently have an altered LOC? (i.e. not alert and calm) • If yes, delirium. • Disorganized brain - Does the patient have a sleep-wake cycle disturbance? (sleeps mostly during the day, has difficulty getting to sleep, does not awaken easily to stimulation, sleeps little at night) • If yes, delirium.

APPENDIX II

USER GUIDE: TRAINING DATASET

Title of dataset: Google Image Search baby image dataset for medical applications

Date of data collection: 2018-12-13

1. Methodological information

This dataset, intended as a resource for training computer vision models, was constructed using a Google Image Search crawler with the following search terms:

- Baby breathing tube,
- Baby eye hospital,
- Baby eyes,
- Baby hospital,
- Baby intubated,
- Baby NG tube.

Annotations were added manually by Vanessa Prinsen using labelImg, an image annotation tool downloaded from Github (Git, 2015).

2. Archive contents

```
#
# ./annotated/
#
```

`./annotated/images/`

665 cleaned and annotated images. There should be no duplicates, extremely low quality images, or images that aren't human (e.g. dolls).

`./annotated/individual annotations/`

Individual annotation files (one for each image).

`./annotated/bounding_boxes.csv`

CSV file containing all annotations (combined from individual annotation files).

`./annotated/check_annotations.py`

Script to check if there are any images without annotations.

`./annotated/process_annotation_files_into_csv.py`

Script to generate one (1) .csv file from all individual .txt and .xml annotation files.

`./annotated/visualize_bounding_boxes.py`

Script to display images with annotated bounding boxes.

```
#
# ./dataset/
#
```

```
./dataset/crawler_baby_breathing_tube/
./dataset/crawler_baby_eye_hospital/
./dataset/crawler_baby_eyes/
./dataset/crawler_baby_hospital/
./dataset/crawler_baby_intubated/
./dataset/crawler_baby_ng_tube/
```

The original images gathered by the Google Image search crawler, separated by search term. There may be duplicates and low-quality images.

```
./dataset/full/
```

All original images gathered together and renamed to avoid conflicts.

3. Data-specific information

.xml annotation files are in Pascal VOC format

.txt annotation files are in the following format:

```
image_id,label,minx,miny,maxx,maxy
```

example: `crawler_baby_breathing_tube_000010.jpg,EYE,34,111,78,157`

APPENDIX III

USER GUIDE: TEST DATASET

Title of dataset: Videos of CHU Sainte-Justine PICU Patients

Date of data collection: 2018-09 to 2019-05

1. Methodological information

This dataset, intended for internal use by researchers at CHU Sainte-Justine, is composed of 59 recordings of patients in the hospital's pediatric intensive care unit. The videos are 5-10 minutes long and were recorded directly in hospital rooms using a standard DSLR photography camera.

Accompanying this dataset are annotation files containing all available bounding boxes coordinates for the eyes and mouth. The 300 image test set used to evaluate eye localization models is also included.

2. Archive contents

```
#
# ./dataset/
#
```

```
./dataset/EYE001/
./dataset/EYE002/
...
./dataset/EYE062/
```

59 folders containing raw recorded video for each patient.

```
./dataset/selected/
```

182 30-second video clips extracted from the full videos (3 for each patient), intended for clinician annotation.

```
./dataset/patient_list_with_ages.xlsx
```

Excel document containing information about each recorded patient: internal identification number, age, date of birth, sex, date of recording, and camera placement during recording.

```
#
# ./test_image_set/
#
```

300 still images extracted randomly from the video recordings to form a test set for eye localization.

```
#
# ./bounding_boxes/
#
```

70 annotation files with the coordinates of the eye and mouth bounding boxes for each frame of recorded video where they could be located.

3. Data-specific information

Bounding box annotation files are provided in a comma-separated (.csv) file, in the following format:

```
frame_number, minx, miny, maxx, maxy, detection
 probability, feature type (eye or mouth)
```

example: `4,825,425,926,516,0.9814,EYE`

APPENDIX IV

IMPLEMENTATION GUIDE: CONVOLUTIONAL NEURAL NETWORKS USING LUMINOTH

This work is accompanied by:

- Tar files containing the models trained for this project
- Modified Luminoth code
- Examples of configuration files used to train the provided models

The basic process for training a convolutional neural network using Luminoth is described below. Instructions are also given for importing existing models. More information about the Luminoth framework can be found at https://luminoth.ai, and full documentation at http://luminoth.readthedocs.io.

1. Organizing the data

Place training data (facial images) in the `./data/train/` directory.
Place the annotations file, train.csv, in the `./data/` directory.
The folder name should match the name of the annotations file.

The first row of the CSV should contain the following headers:

```
image_id, label, xmin, ymin, xmax, ymax
```

This is followed by any number of data rows:

```
10.jpg, MOUTH, 59, 167, 144, 210
10.jpg, EYE, 111, 77, 162, 128
10.jpg, MOUTH, 59, 167, 144, 210
```

2. Transforming the data

This changes the data into a format usable by Luminoth. Multiple input types are supported, but we have used csv files for our work.

```
lumi dataset transform --type csv  --data-dir ./data/
 --output-dir ./transformed_images/ --split train
```

`--type`: type of annotation file

`--data-dir`: directory containing input annotation file and image folder

`--output-dir`: where to place the transformed data (.tfrecords file)

`--split`: name of the annotation file/data folder to use

3. Training the model

```
lumi train -c train50_config.yml
```

Sample yml configuration files are provided with the Luminoth toolkit. The following is an example of a configuration file, `train50_config.yml`, that will train a convolutional neural network for 50 epochs, using a Faster R-CNN architecture and the Resnet-101 base network:

```
train:
  # Run name for the training session.
  run_name: train50-run
  # Directory in which model checkpoints & summaries (for Tensorboard)
will be saved.
  job_dir: jobs/
  # Number of epochs (complete dataset batches) to run.
  num_epochs: 50
```

```
dataset:
  type: object_detection
  # From which directory to read the dataset.
  dir: ’transformed_images/’
  # Which split of tfrecords to look for.
  split: train

model:
  type: fasterrcnn
  network:
    # Total number of classes to predict.
    num_classes: 2
  base_network:
    architecture: resnet_v1_101
    # From which path to load the weights
    weights_path: /home/downloaded_weights/
    # Should we download weights if not available.
    download: False
```

The Luminoth code used for our project was modified to support the final two options, which allow the base network and weights to be downloaded beforehand and stored locally at a specified location. This allowed us to train our model on a computer without external internet access.

If these options are left out, Luminoth will attempt to download the base network and weights from the Internet if it cannot find them in its default cache location.

4. Predicting using Luminoth

```
lumi predict image.png
```

A Jupyter notebook, "Visualize with Luminoth.ipynb", is provided as an example of how to use trained Luminoth models with Python.

5. Saving and loading trained models

Luminoth allows saving models as "checkpoints". These can be exported to a tar file and reimported later.

To create a checkpoint from a trained model:

```
lumi checkpoint create train50_config.yml -e name="Sample
   trained model (50 ep)" -e alias=train50
```

To view a list of checkpoints:

```
lumi checkpoint list
```

To export a checkpoint (generates a tar file):

```
lumi checkpoint export <checkpoint id from list command>
```

To import a checkpoint from a tar file:

```
lumi checkpoint import <tar file name>
```

REFERENCES

Abbood, H., Al-Nuaimy, W., Al-Ataby, A., Salem, S. A. & AlZubi, H. S. (2014). *Prediction of driver fatigue: Approaches and open challenges*. Conference Proceedings presented in 2014 14th UK Workshop on Computational Intelligence (UKCI) (pp. 1-6). doi: 10.1109/UKCI.2014.6930193.

Adachi, K., Yamamto, N., Yamamoto, O., Nakano, T. & Yamamoto, S. (2006). Monitoring car drivers' condition using image processing - measurement of car drivers' consciousness in consideration of individual differences. *Transactions of the Institute of Electrical Engineers of Japan, Part E*, 126-E(2), 31-7. doi: 10.1541/ieejsmas.126.31.

Akrout, B. & Mahdi, W. (2013). *A visual based approach for drowsiness detection*. Conference Proceedings presented in 2013 IEEE Intelligent Vehicles Symposium (IV) (pp. 1324-1329). doi: 10.1109/IVS.2013.6629650.

Akrout, B. & Mahdi, W. (2015). Spatio-temporal features for the automatic control of driver drowsiness state and lack of concentration. *Machine Vision and Applications*, 26(1), 1-13. doi: 10.1007/s00138-014-0644-z.

Al-Moteri, M. O., Symmons, M., Plummer, V. & Cooper, S. (2017). Eye tracking to investigate cue processing in medical decision-making: A scoping review. *Computers in Human Behavior*, 66, 52-66. doi: 10.1016/j.chb.2016.09.022.

Al-Rahayfeh, A. & Faezipour, M. (2014). *Classifiers comparison for a new eye gaze direction classification system*. Conference Proceedings presented in 2014 IEEE Long Island Systems, Applications and Technology Conference (LISAT), 2 May 2014 (pp. 6 pp.). doi: 10.1109/LISAT.2014.6845189.

Albu, A. B., Widsten, B., Tiange, W., Lan, J. & Mah, J. (2008). *A computer vision-based system for real-time detection of sleep onset in fatigued drivers*. Conference Proceedings presented in 2008 IEEE Intelligent Vehicles Symposium (IV), 4-6 June 2008 (pp. 25-30). doi: 10.1109/IVS.2008.4621133.

Ambuel, B., Hamlett, K. W., Marx, C. M. & Blumer, J. L. (1992). Assessing distress in pediatric intensive care environments: the COMFORT scale. *Journal of pediatric psychology*, 17(1), 95-109. doi: 10.1093/jpepsy/17.1.95.

Amiel-Tison, C., Barrier, G., Shnider, S. M., Levinson, G., Hughes, S. C. & Stefani, S. J. (1982). A new neurologic and adaptive capacity scoring system for evaluating obstetric medications in full-term newborns. *Anesthesiology*, 56(5), 340-350. doi: 10.1097/00000542-198205000-00003.

Arbour, C., ChoiniÃšre, M., Topolovec-Vranic, J., Loiselle, C. G., Puntillo, K. & GÃ©linas, C. (2014). Detecting pain in traumatic brain-injured patients with different levels of consciousness during common procedures in the ICU: typical or atypical behaviors? *The Clinical journal of pain*, 30(11), 960-969. doi: 10.1097/AJP.0000000000000061.

Ashraf, A. B., Lucey, S., Cohn, J. F., Chen, T., Ambadar, Z., Prkachin, K. M. & Solomon, P. E. (2009). The painful face–pain expression recognition using active appearance models. *Image and vision computing*, 27(12), 1788-1796. doi: 10.1016/j.imavis.2009.05.007.

Aslin, R. N. (2012). Infant Eyes: A Window on Cognitive Development. *Infancy*, 17(1), 126-140. doi: doi:10.1111/j.1532-7078.2011.00097.x.

Aung, P. W. A., Foo, S. F. V., Huang, W., Biswas, J., Phua, J. E., Liou, K. & Hsia, C.-C. (2010). *Evaluation and analysis of multimodal sensors for developing in and around the bed patient monitoring system*. Conference Proceedings presented in 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology (pp. 2159-2162).

Avner, J. R. (2006). Altered states of consciousness. *Pediatrics in review*, 27(9), 331-338. doi: 10.1542/pir.27-9-331.

Bainbridge, W. A., Isola, P. & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General*, 142(4), 1323. doi: 10.1037/a0033872.

Baltrusaitis, T., Zadeh, A., Lim, Y. C. & Morency, L.-P. (2018). *OpenFace 2.0: Facial Behavior Analysis Toolkit*. Conference Proceedings presented in Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on (pp. 59-66).

Ban, Y., Kim, S.-K., Kim, S., Toh, K.-A. & Lee, S. (2014). Face detection based on skin color likelihood. *Pattern Recognition*, 47(4), 1573-1585. doi: https://doi.org/10.1016/j.patcog.2013.11.005.

Bauer, P., Kramer, J. B., Rush, B. & Sabalka, L. (2017). *Modeling bed exit likelihood in a camera-based automated video monitoring application*. Conference Proceedings presented in 2017 IEEE International Conference on Electro Information Technology, EIT 2017, May 14, 2017 - May 17, 2017 (pp. 56-61). doi: 10.1109/EIT.2017.8053330.

Bazrafkan, S., Kar, A. & Costache, C. (2015). Eye Gaze for Consumer Electronics: Controlling and commanding intelligent systems. *IEEE Consumer Electronics Magazine*, 4(4), 65-71. doi: 10.1109/MCE.2015.2464852.

Berstein, J., Risemberg, H. M., Dorst, J. P., Rowe, R. D. & Spear, G. S. (1972). The respiratory distress syndrome and/or cyanotic congenital heart disease. *The Journal of Pediatrics*, 81(3), 609-616. doi: 10.1016/S0022-3476(72)80200-2.

BioID. (n.d.). BioID Face Database [Dataset]. Consulted at https://www.bioid.com/facedb/.

Boerlage, A., Ista, E., Duivenvoorden, H., De Wildt, S., Tibboel, D. & van Dijk, M. (2015). The COMFORT behaviour scale detects clinically meaningful effects of analgesic and sedative treatment. *European Journal of Pain*, 19(4), 473-479. doi: 10.1002/ejp.569.

Borys, M., Plechawska-Wójcik, M., Wawrzyk, M. & Wesołowska, K. (2017). Classifying cognitive workload using eye activity and eeg features in arithmetic tasks. In *Communications in Computer and Information Science* (vol. 756, pp. 90-105). doi: 10.1007/978-3-319-67642-5_8.

Bradski, G. & Kaehler, A. (2000). OpenCV. *Dr. Dobb's journal of software tools*, 25(11), 120-+. Consulted at <GotoISI>://WOS:000089917100022.

Brahnam, S., Chuang, C.-F., Shih, F. Y. & Slack, M. R. (2006). Machine recognition and representation of neonatal facial displays of acute pain. *Artificial Intelligence in Medicine*, 36(3), 211-222. doi: https://doi.org/10.1016/j.artmed.2004.12.003.

Brahnam, S., Chuang, C.-F., Sexton, R. S. & Shih, F. Y. (2007a). Machine assessment of neonatal facial expressions of acute pain. *Decision Support Systems*, 43(4), 1242-1254. doi: 10.1016/j.dss.2006.02.004.

Brahnam, S., Nanni, L. & Sexton, R. (2007b). Introduction to neonatal facial pain detection using common and advanced face classification techniques. In *Advanced Computational Intelligence Paradigms in Healthcare* (pp. 225-253). Springer.

Brand, J. & Mason, J. S. (2000). *A comparative assessment of three approaches to pixel-level human skin-detection*. Conference Proceedings presented in wecwis (pp. 5056).

Chai, D. & Bouzerdoum, A. (2000). *A Bayesian approach to skin color classification in YCbCr color space*. Conference Proceedings presented in TENCON 2000. Proceedings (pp. 421-424).

Chaichulee, S., Villarroel, M., Jorge, J., Arteta, C., Green, G., McCormick, K., Zisserman, A. & Tarassenko, L. (2017). *Multi-task convolutional neural network for patient detection and skin segmentation in continuous non-contact vital sign monitoring*. Conference Proceedings presented in 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) (pp. 266-272).

Chase, J. G., Agogue, F., Starfinger, C., Lam, Z., Shaw, G. M., Rudge, A. D. & Sirisena, H. (2004a). Quantifying agitation in sedated ICU patients using digital imaging. *Computer methods and programs in biomedicine*, 76(2), 131-141. doi: 10.1016/j.cmpb.2004.03.005.

Chase, J. G., Agogue, F., Starfinger, C., Lam, Z., Shaw, G. M., Rudge, A. D. & Sirisena, H. (2004b). Quantifying agitation in sedated ICU patients using digital imaging. *Computer methods and programs in biomedicine*, 76(2), 131-141. doi: 10.1016/j.cmpb.2004.03.005.

Chen, K., Gabriel, P., Alasfour, A., Gong, C., Doyle, W. K., Devinsky, O., Friedman, D., Dugan, P., Melloni, L. & Thesen, T. (2018). Patient-Specific Pose Estimation in Clinical Environments. *IEEE Journal of Translational Engineering in Health and Medicine*, 6, 1-11. doi: 10.1109/JTEHM.2018.2875464.

Chen, K. J. (2018). *Continuous Behavior Acquisition in Clinical Environments*. (Thesis).

Chen, S. & Epps, J. (2013). Automatic classification of eye activity for cognitive load measurement with emotion interference. *Computer Methods and Programs in Biomedicine*, 110(2), 111-24. doi: 10.1016/j.cmpb.2012.10.021.

Chen, S., Epps, J. & Chen, F. (2013). *Automatic and continuous user task analysis via eye activity*. Conference Proceedings presented in International Conference on Intelligent User Interfaces, Proceedings IUI (pp. 57-66). doi: 10.1145/2449396.2449406.

Chen, S. & Liu, C. (2011). *Fast eye detection using different color spaces*. Conference Proceedings presented in Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on (pp. 521-526).

Chen, S., Epps, J., Ruiz, N. & Chen, F. (2011). *Eye activity as a measure of human mental effort in HCI*. Conference Proceedings presented in 2011 15th ACM International Conference on Intelligent User Interfaces, IUI'11, February 13, 2011 - February 16, 2011 (pp. 315-318). doi: 10.1145/1943403.1943454.

Children's Hospital of Philadelphia. (n.d.). Developmental anchors for CAPD [Journal Article]. Consulted at http://www.chop.edu/clinical-pathway/picu-pcu-delirium-developmental-anchors.

Ching-Wei, W. & Hunter, A. (2008a). *A robust pose matching algorithm for covered body analysis for sleep apnea*. Conference Proceedings presented in 2008 8th IEEE International Conference on Bioinformatics and BioEngineering, 8-10 Oct. 2008 (pp. 7 pp.). doi: 10.1109/BIBE.2008.4696847.

Ching-Wei, W. & Hunter, A. (2008b). A novel approach to detect the obscured upper body in application to diagnosis of obstructive sleep apnea. *IAENG International Journal of Computer Science*, 35(1), 110-18.

Chouinard, B., Scott, K. & Cusack, R. (2019). Using automatic face analysis to score infant behaviour from video collected online. *Infant Behavior and Development*, 54, 1-12.

doi: https://doi.org/10.1016/j.infbeh.2018.11.004.

Chuanxiang, T., Jiwu, L. & Jie, L. (2018). *Non-contact Heart Rate Monitoring by Combining Convolutional Neural Network Skin Detection and Remote Photoplethysmography via a Low-cost Camera*. Conference Proceedings presented in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 18-22 June 2018 (pp. 1390-6). doi: 10.1109/CVPRW.2018.00178.

Chun-Ning, M., Tai-Ning, Z., Pin, Z. & Sheng-Jiang, C. (2012). Fast and precise iris localization for low-resolution facial images. *Optical Engineering*, 51(7), 077008 (10 pp.). doi: 10.1117/1.OE.51.7.077008.

Chwyl, B., Chung, A. G., Shafiee, M. J., Fu, Y. & Wong, A. (2017). *DeepPredict: A deep predictive intelligence platform for patient monitoring*. Conference Proceedings presented in 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2017, July 11, 2017 - July 15, 2017 (pp. 4309-4312). doi: 10.1109/EMBC.2017.8037809.

Corbetta, D., Guan, Y. & Williams, J. L. (2012). Infant Eye-Tracking in the Context of Goal-Directed Actions. *Infancy*, 17(1), 102-125. doi: doi:10.1111/j.1532-7078.2011.00093.x.

Dalal, N. & Triggs, B. (2005). *Histograms of oriented gradients for human detection*. Conference Proceedings presented in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) (pp. 886-893 vol. 1). doi: 10.1109/CVPR.2005.177.

Danielsen, A. (2016). *Non-intrusive bedside event recognition using infrared array and ultrasonic sensor*. Conference Proceedings presented in 10th International Conference on Ubiquitous Computing and Ambient Intelligence, UCAmI 2016, November 29, 2016 - December 2, 2016 (pp. 15-25). doi: 10.1007/978-3-319-48746-5_2.

Danisman, T., Bilasco, I. M., Djeraba, C. & Ihaddadene, N. (2010). *Drowsy driver detection system using eye blink patterns*. Conference Proceedings presented in 2010 International Conference on Machine and Web Intelligence (pp. 230-233). doi: 10.1109/ICMWI.2010.5648121.

Davidson, A. J. (2007). The aims of anesthesia in infants: the relevance of philosophy, psychology and a little evidence. *Pediatric Anesthesia*, 17(2), 102-108. doi: 10.1111/j.1460-9592.2006.02053.x.

De Rivecourt, M., Kuperus, M. N., Post, W. J. & Mulder, L. J. M. (2008). Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics*, 51(9), 1295-1319. doi: 10.1080/00140130802120267.

Dell, N., Francis, I., Sheppard, H., Simbi, R. & Borriello, G. (2014). *Field evaluation of a camera-based mobile health system in low-resource settings*. Conference Proceedings presented in 16th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2014, September 23, 2014 - September 26, 2014 (pp. 33-42). doi: 10.1145/2628363.2628366.

Duncan, C. C., Ment, L. R., Smith, B. & Ehrenkranz, R. A. (1981). A scale for the assessment of neonatal neurologic status. *Pediatric Neurosurgery*, 8(4), 299-306. doi: 10.1159/000119993.

DURHAM, S. R., CLANCY, R. R., LEUTHARDT, E., SUN, P., KAMERLING, S., DOMINGUEZ, T. & DUHAIME, A.-C. (2000). CHOP Infant Coma Scale (" Infant Face Scale"): a novel coma scale for children less than two years of age. *Journal of neurotrauma*, 17(9), 729-737. doi: DOI 10.1089/neu.2000.17.729.

Dwivedi, K., Biswaranjan, K. & Sethi, A. (2014). *Drowsy driver detection using representation learning*. Conference Proceedings presented in 2014 IEEE International Advance Computing Conference (IACC) (pp. 995-999).

Eastwood-Sutherland, C., Gale, T. J., Dargavillc, P. A. & Wheeler, K. (2016). *Elements of vision based respiratory monitoring*. Conference Proceedings presented in 8th Biomedical Engineering International Conference, BMEiCON 2015, November 25, 2015 - November 27, 2015. doi: 10.1109/BMEiCON.2015.7399536.

Edlow, B. L., Chatelle, C., Spencer, C. A., Chu, C. J., Bodien, Y. G., O'Connor, K. L., Hirschberg, R. E., Hochberg, L. R., Giacino, J. T. & Rosenthal, E. S. (2017). Early detection of consciousness in patients with acute severe traumatic brain injury. *Brain*, 140(9), 2399-2414. doi: 10.1093/brain/awx176.

Ekman, P. & Friesen, W. V. (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior*, 1(1), 56-75. doi: 10.1007/bf01115465.

El Kaddouhi, S., Saaidi, A. & Abarkan, M. (2017). Eye detection based on the Viola-Jones method and corners points. *Multimedia Tools and Applications*, 76(21), 23077-97. doi: 10.1007/s11042-017-4415-5.

Elsharkawi, N. G. (1999). Simple pediatric analog sedation score (PASS). *Anesthesia & Analgesia*, 88(1), 227. doi: 10.1097/00000539-199901000-00043.

Emmanuel, J. A. & Ibiyemi, T. S. (2017). A Hybrid Skin Colour Based Face Detection System to Optimally Detect Black Faces. *American Journal of Intelligent Systems*, 7(4), 120-124.

Fasel, I., Fortenberry, B. & Movellan, J. (2005). A generative framework for real time object detection and classification. *Computer Vision and Image Understanding*, 98(1), 182-210. doi: 10.1016/j.cviu.2004.07.014.

Fei, Z. & Qiang, Q. (2009). *Face detection based on rectangular knowledge rule and face structure*. Conference Proceedings presented in 2009 First International Conference on Information Science and Engineering (pp. 1235-1239).

Feng, G. C. & Yuen, P. C. (1998). Variance projection function and its application to eye detection for human face recognition. *Pattern Recognition Letters*, 19(9), 899-906. doi: https://doi.org/10.1016/S0167-8655(98)00065-8.

Fotiadou, E., Zinger, S., Ten, W. E. T. a., Oetomo, S. B. & With, P. H. N. d. (2014). *Video-based facial discomfort analysis for infants*. Conference Proceedings presented in IS&T/SPIE Electronic Imaging (pp. 14).

Fuangkaew, S. & Patanukhom, K. (2013). *Eye State Detection and Eye Sequence Classification for Paralyzed Patient Interaction*. Conference Proceedings presented in 2013 2nd IAPR Asian Conference on Pattern Recognition (ACPR), 5-8 Nov. 2013 (pp. 376-80). doi: 10.1109/ACPR.2013.91.

Gambi, E., Agostinelli, A., Belli, A., Burattini, L., Cippitelli, E., Fioretti, S., Pierleoni, P., Ricciuti, M., Sbrollini, A. & Spinsante, S. (2017). Heart Rate Detection Using Microsoft Kinect: Validation and Comparison to Wearable Devices. *Sensors*, 17(8), 1776. doi: ARTN 1776 10.3390/s17081776.

Garcia-Garcia, M., Caplier, A. & Rombaut, M. (2018). *Sleep deprivation detection for real-time driver monitoring using deep learning*. Conference Proceedings presented in Image Analysis and Recognition. 15th International Conference, ICIAR 2018, 27-29 June 2018 (pp. 435-42). doi: 10.1007/978-3-319-93000-8_49.

Ge, W., Han, C. & Quan, W. (2015). *Face detection in complex background based on Adaboost algorithm and YCbCr skin color model*. Conference Proceedings presented in MIPPR 2015: Pattern Recognition and Computer Vision (pp. 981312).

Ghazali, K. H. B., Ma, J. & Xiao, R. (2011). An innovative face detection based on skin color segmentation. *International Journal of Computer Applications*, 34(2), 6-10.

Gholami, B., Haddad, W. M. & Tannenbaum, A. R. (2009). *Agitation and pain assessment using digital imaging*. Conference Proceedings presented in Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE (pp. 2176-2179).

GitHub. (2015). labelImg [Computer Program]. Consulted at https://github.com/tzutalin/labelImg.

Gou, C., Wu, Y., Wang, K., Wang, K., Wang, F.-Y. & Ji, Q. (2017). A joint cascaded framework for simultaneous eye detection and eye state estimation. *Pattern Recognition*, 67, 23-31. doi: 10.1016/j.patcog.2017.01.023.

Grech, V. & Goldman, A. (2001). Acute respiratory distress syndrome due to cyanotic spell in an infant with tetralogy of fallot. *Pediatric Pulmonology*, 32(5), 406-407. doi: 10.1002/ppul.1151.

Gross, R. (2005). Face Databases. In Li, S. Z. & Jain, A. K. (Eds.), *Handbook of Face Recognition* (ch. Chapter 14, pp. 301-327). New York, NY: Springer New York. doi: 10.1007/0-387-27257-7_14.

Grunau, R. V. & Craig, K. D. (1987). Pain expression in neonates: facial action and cry. *Pain*, 28(3), 395-410. doi: 10.1016/0304-3959(87)90073-x.

Gupta, I., Garg, N., Aggarwal, A., Nepalia, N. & Verma, B. (2018). *Real-Time Driver's Drowsiness Monitoring Based on Dynamically Varying Threshold*. Conference Proceedings presented in 2018 Eleventh International Conference on Contemporary Computing (IC3) (pp. 1-6). doi: 10.1109/IC3.2018.8530651.

Hahn, Y. S., Chyung, C., Barthel, M. J., Bailes, J., Flannery, A. M. & McLone, D. G. (1988). Head injuries in children under 36 months of age. *Child's Nervous System*, 4(1), 34-39. doi: 10.1007/bf00274081.

Han, W., Yang, Y., Huang, G. B., Sourina, O., Klanner, F. & Denk, C. (2015). *Driver Drowsiness Detection Based on Novel Eye Openness Recognition Method and Unsupervised Feature Learning*. Conference Proceedings presented in 2015 IEEE International Conference on Systems, Man, and Cybernetics (pp. 1470-1475). doi: 10.1109/SMC.2015.260.

Harezlak, K., Kasprowski, P., Dzierzega, M. & Kruk, K. (2016). *Application of Eye Tracking for Diagnosis and Therapy of Children with Brain Disabilities* (pp. 323-333).

He, K., Zhang, X., Ren, S. & Sun, J. (2016). *Deep residual learning for image recognition*. Conference Proceedings presented in Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Hoffman, G. M., Nowakowski, R., Troshynski, T. J., Berens, R. J. & Weisman, S. J. (2002). Risk reduction in pediatric procedural sedation by application of an American Academy of Pediatrics/American Society of Anesthesiologists process model. *Pediatrics*, 109(2), 236-243. doi: 10.1542/peds.109.2.236.

Hsu, R.-L., Abdel-Mottaleb, M. & Jain, A. K. (2002). Face detection in color images. *IEEE transactions on pattern analysis and machine intelligence*, 24(5), 696-706. doi: Doi 10.1109/34.1000242.

Hu, M., Zhai, G., Li, D., Fan, Y., Duan, H., Zhu, W. & Yang, X. (2018). Combination of near-infrared and thermal imaging techniques for the remote and simultaneous measurements of breathing and heart rates under sleep situation. *PLoS One*, 13(1), e0190466. doi: 10.1371/journal.pone.0190466.

Ilias, T., Filip, B., Radu, C., Dag, N., Marina, S. & Mevludin, M. (2017). *Using measurements from wearable sensors for automatic scoring of Parkinson's disease motor states: Results from 7 patients*. Conference Proceedings presented in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 131-134). doi: 10.1109/EMBC.2017.8036779.

Inoue, M., Taguchi, R. & Umezaki, T. (2018a). *Vision-based Bed Detection for Hospital Patient Monitoring System*. Conference Proceedings presented in 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2018, July 18, 2018 - July 21, 2018 (pp. 5006-5009). doi: 10.1109/EMBC.2018.8513460.

Inoue, M., Yasui, T., Taguchi, R. & Umezaki, T. (2018b). Detection of bed position for automation of patient monitoring function. *IEEJ Transactions on Electronics, Information and Systems*, 138(6), 670-677. doi: 10.1541/ieejeiss.138.670.

Irani, R., Nasrollahi, K. & Moeslund, T. B. (2014a). *Pain recognition using spatiotemporal oriented energy of facial muscles*. Conference Proceedings presented in IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2015, June 7, 2015 - June 12, 2015 (pp. 80-87). doi: 10.1109/CVPRW.2015.7301340.

Irani, R., Nasrollahi, K., Simon, M. O., Corneanu, C. A., Escalera, S., Bahnsen, C., Lundtoft, D. H., Moeslund, T. B., Pedersen, T. L., Klitgaard, M.-L. & Petrini, L. (2014b). *Spatiotemporal analysis of RGB-D-T facial images for multimodal pain level recognition*. Conference Proceedings presented in IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2015, June 7, 2015 - June 12, 2015 (pp. 88-95). doi: 10.1109/CVPRW.2015.7301341.

Ista, E., van Dijk, M., Tibboel, D. & de Hoog, M. (2005). Assessment of sedation levels in pediatric intensive care patients can be improved by using the COMFORT "behavior" scale*. *Pediatric Critical Care Medicine*, 6(1), 58-63. doi: 10.1097/01.Pcc.0000149318.40279.1a.

Jang, Y.-M., Mallipeddi, R., Lee, S., Kwak, H.-W. & Lee, M. (2014). Human intention recognition based on eyeball movement pattern and pupil size variation. *Neurocomputing*, 128, 421-432. doi: 10.1016/j.neucom.2013.08.008.

Jeena, P. M. (2008). An approach to the child in respiratory distress. *South African Family Practice*, 50(3), 32-37. Consulted at https://www.scopus.com/inward/record.uri?eid=2-s2.0-60249085931&partnerID=40&md5=e799618b13d09fcc076a5c5947ce32d9.

Jesorsky, O., Kirchberg, K. J. & Frischholz, R. W. (2001). *Robust Face Detection Using the Hausdorff Distance*. Conference Proceedings presented in Audio- and Video-Based Biometric Person Authentication (pp. 90-95).

Ji, Y., Wang, S., Zhao, Y., Wei, J. & Lu, Y. (2019). Fatigue State Detection Based on Multi-Index Fusion and State Recognition Network. *IEEE Access*, 7, 64136-64147. doi: 10.1109/ACCESS.2019.2917382.

Jo, J., Lee, S. J., Park, K. R., Kim, I.-J. & Kim, J. (2014). Detecting driver drowsiness using feature-level fusion and user-specific classification. *Expert Systems with Applications*, 41(4, Part 1), 1139-1152. doi: https://doi.org/10.1016/j.eswa.2013.07.108.

Johansson, M. & Kokinsky, E. (2009). The COMFORT behavioural scale and the modified FLACC scale in paediatric intensive care. *Nursing in critical care*, 14(3), 122-130. doi: 10.1111/j.1478-5153.2009.00323.x.

Jurkowlaniec, A., Dybizbanski, T., Szulc, M., Figas, A., Gugala, K., Rybarczyk, A. & Michalak, S. (2011). *The system supporting the clinical observation of eyeball movements*. Conference Proceedings presented in 18th International Conference - Mixed Design of Integrated Circuits and Systems, MIXDES 2011, June 16, 2011 - June 18, 2011 (pp. 667-672).

Just, M. A. & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, 8(4), 441-480. doi: https://doi.org/10.1016/0010-0285(76)90015-3.

Kakumanu, P., Makrogiannis, S. & Bourbakis, N. (2007). A survey of skin-color modeling and detection methods. *Pattern recognition*, 40(3), 1106-1122. doi: 10.1016/j.patcog.2006.06.010.

Kaur, M., Marshall, A. P., Eastwood-Sutherland, C., Salmon, B. P., Dargaville, P. A. & Gale, T. J. (2017). Automatic torso detection in images of preterm infants. *Journal of medical systems*, 41(9), 134. doi: 10.1007/s10916-017-0782-8.

Khairosfaizal, W. M. K. W. M. & Nor'aini, A. J. (2009). *Eyes detection in facial images using Circular Hough Transform*. Conference Proceedings presented in 2009 5th International Colloquium on Signal Processing & Its Applications (pp. 238-242). doi: 10.1109/CSPA.2009.5069224.

Khan, R. A., Meyer, A., Konik, H. & Bouakaz, S. (2013). *Pain detection through shape and appearance features*. Conference Proceedings presented in 2013 IEEE International

Conference on Multimedia and Expo, ICME 2013, July 15, 2013 - July 19, 2013. doi: 10.1109/ICME.2013.6607608.

Khosla, A., Bainbridge, W. A., Torralba, A. & Oliva, A. (2013). *The intrinsic memorability of face photographs*. Conference Proceedings presented in In Journal of Experimental Psychology: General.

Kirana, M. C., Purnama, I. K. E., Suprapto, Y. K., Hariadi, M. & Purnomo, M. H. (2013). *Facial feature extraction on pre and post-operative infant with NFCS and nCRF*. Conference Proceedings presented in 2013 3rd International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME), 7-8 Nov. 2013 (pp. 54-8). doi: 10.1109/ICICI-BME.2013.6698464.

Kirkham, F. J., Newton, C. R. & Whitehouse, W. (2008). Paediatric coma scales. *Developmental medicine & child neurology*, 50(4), 267-274. doi: 10.1111/j.1469-8749.2008.02042.x.

Kittipanya-Ngam, P., Ong Soh, G. & Eng How, L. (2012a). *Bed detection for monitoring system in hospital wards*. Conference Proceedings presented in 2012 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 28 Aug.-1 Sept. 2012 (pp. 5887-90). doi: 10.1109/EMBC.2012.6347333.

Kittipanya-Ngam, P., Guat, O. S. & Lung, E. H. (2012b). *Computer vision applications for patients monitoring system*. Conference Proceedings presented in 15th International Conference on Information Fusion, FUSION 2012, September 7, 2012 - September 12, 2012 (pp. 2201-2208).

Kojima, N., Kozuka, K., Nakano, T. & Yamamoto, S. (2001). *Detection of consciousness degradation and concentration of a driver for friendly information service*. Conference Proceedings presented in IVEC2001. Proceedings of the IEEE International Vehicle Electronics Conference 2001. IVEC 2001 (Cat. No.01EX522) (pp. 31-36). doi: 10.1109/IVEC.2001.961722.

Kolkur, S., Kalbande, D., Shimpi, P., Bapat, C. & Jatakia, J. (2017). Human skin detection using RGB, HSV and YCbCr color models. *arXiv preprint arXiv:1708.02694*, 137, 324-332. Consulted at <GotoISI>://WOS:000417222900051.

Kotropoulos, C. & Pitas, I. (1997). *Rule-based face detection in frontal views*. Conference Proceedings presented in Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on (pp. 2537-2540).

Kovac, J., Peer, P. & Solina, F. (2003). *Human skin color clustering for face detection*. Conference Proceedings presented in The IEEE Region 8 EUROCON 2003. Computer as a Tool. (pp. 144-148 vol.2). doi: 10.1109/EURCON.2003.1248169.

Krieger, H. P., Wagman, I. H. & Bender, M. B. (1958). CHANGES IN STATE OF CONSCIOUSNESS AND PATTERNS OF EYE MOVEMENTS. *Journal of Neurophysiology*, 21(3), 224-230. doi: 10.1152/jn.1958.21.3.224.

Krizhevsky, A. & Hinton, G. (2009). *Learning multiple layers of features from tiny images*.

Learn OpenCV. (2016). Histogram of Oriented Gradients [Web Page]. Consulted at https://www.learnopencv.com/histogram-of-oriented-gradients/.

Lebreton, P., Hupont, I., Hirth, M., MÃ€ki, T., Skodras, E., Schubert, A. & Raake, A. (2019). CrowdWatcher: an open-source platform to catch the eye of the crowd. *Quality and User Experience*, 4(1), 1.

Lee, G., Ojha, A. & Lee, M. (2015a). *Concentration monitoring for intelligent tutoring system based on pupil and eye-blink*. Conference Proceedings presented in HAI 2015 - Proceedings of the 3rd International Conference on Human-Agent Interaction (pp. 291-294). doi: 10.1145/2814940.2815000.

Lee, G., Ojha, A. & Lee, M. (2015b). *Classification of High and Low Intelligent Individuals Using Pupil and Eye Blink*. Conference Proceedings presented in International Conference on Neural Information Processing (pp. 459-466).

Lee, K.-B., Kim, D.-J. & Hong, K.-S. (2011). *An Implementation of SVM-Based Gaze Recognition System Using Advanced Eye Region Detection*. Conference Proceedings presented in Computational Science and Its Applications - ICCSA 2011 (pp. 58-69).

Lee, K. H., Kim, W., Choi, H. K. & Jang, B. T. (2019). *A Study on Feature Extraction Methods Used to Estimate a Driver's Level of Drowsiness*. Conference Proceedings presented in 2019 21st International Conference on Advanced Communication Technology (ICACT) (pp. 710-713).

Lempert, T. & Von Brevern, M. (1996). The eye movements of syncope. *Neurology*, 46(4), 1086-1088. doi: 10.1212/wnl.46.4.1086.

Lester, B. M. & Tronick, E. Z. (2004a). The neonatal intensive care unit network neurobehavioral scale procedures. *Pediatrics*, 113(Supplement 2), 641-667. Consulted at https://www.ncbi.nlm.nih.gov/pubmed/14993524.

Lester, B. M. & Tronick, E. Z. (2004b). The neonatal intensive care unit network neurobehavioral scale procedures. *Pediatrics*, 113(Supplement 2), 641-667. Consulted at https://www.ncbi.nlm.nih.gov/pubmed/14993524.

Li, C., Zinger, S., a Ten, W. T. & de With, P. (2016). *Video-based discomfort detection for infants using a Constrained Local Model*. Conference Proceedings presented in 2016 International Conference on Systems, Signals and Image Processing (IWSSIP) (pp. 1-4).

Liu, J., Wang, W., Liu, F. & Li, Z. (2018). Pediatric acute respiratory distress syndrome - Current views (Review). *Experimental and Therapeutic Medicine*, 15(2), 1775-1780. doi: 10.3892/etm.2017.5628.

Liu, S. & Ostadabbas, S. (2018). *A Vision-Based System for In-Bed Posture Tracking*. Conference Proceedings presented in 16th IEEE International Conference on Computer Vision Workshops, ICCVW 2017, October 22, 2017 - October 29, 2017 (pp. 1373-1382). doi: 10.1109/ICCVW.2017.163.

Liu, S., Yin, Y. & Ostadabbas, S. (2019a). In-Bed Pose Estimation: Deep Learning with Shallow Dataset. *IEEE Journal of Translational Engineering in Health and Medicine*, 7, 4900112. doi: 10.1109/JTEHM.2019.2892970.

Liu, W., Qian, J., Yao, Z., Jiao, X. & Pan, J. (2019b). Convolutional Two-Stream Network Using Multi-Facial Feature Fusion for Driver Fatigue Detection. *Future Internet*, 11(5), 115. doi: ARTN 115 10.3390/fi11050115.

Liu, Y., Wu, Q., Tang, L. & Shi, H. (2017). Gaze-Assisted Multi-Stream Deep Neural Network for Action Recognition. *IEEE Access*, 5, 19432-19441. doi: 10.1109/ACCESS.2017.2753830.

Lopar, M. & Ribarić, S. (2013). An Overview and Evaluation of Various Face and Eyes Detection Algorithms for Driver Fatigue Monitoring Systems. *arXiv preprint arXiv:1310.0317*.

Luo, S.-Y. (2015). *Design and Implementation of Remote Medical Nursing Monitoring System Based on Computer Network*. Conference Proceedings presented in 7th International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2015, June 13, 2015 - June 14, 2015 (pp. 346-349). doi: 10.1109/ICMTMA.2015.89.

Lv, C., Zhang, T. & Lin, C. (2017). *Face detection based on skin color and AdaBoost algorithm*. Conference Proceedings presented in Control And Decision Conference (CCDC), 2017 29th Chinese (pp. 1363-1367).

Lyu, J., Yuan, Z. & Chen, D. (2018). Long-term multi-granularity deep framework for driver drowsiness detection. *arXiv preprint arXiv:1801.02325*.

Macnab, A. J., Levine, M., Glick, N., Phillips, N., Susak, L. & Elliott, M. (1994a). The Vancouver sedative recovery scale for children: validation and reliability of scoring based on videotaped instruction. *Canadian journal of anaesthesia*, 41(10), 913. doi: 10.1007/BF03010934.

Macnab, A. J., Levine, M., Glick, N., Phillips, N., Susak, L. & Elliott, M. (1994b). The Vancouver sedative recovery scale for children: validation and reliability of scoring based on videotaped instruction. *Canadian journal of anaesthesia*, 41(10), 913.

Mahadevi, M. & Sumathi, C. (2015). Face Localization based on Skin Color. *International Journal of Computer Applications*, 109(12), 25-28.

Malviya, S., Voepel-Lewis, T., Tait, A., Merkel, S., Tremper, K. & Naughton, N. (2002). Depth of sedation in children undergoing computed tomography: validity and reliability of the University of Michigan Sedation Scale (UMSS). *British journal of anaesthesia*, 88(2), 241-245. doi: 10.1093/bja/88.2.241.

Mandal, B., Li, L., Wang, G. S. & Lin, J. (2017). Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State. *IEEE Transactions on Intelligent Transportation Systems*, 18(3), 545-557. doi: 10.1109/TITS.2016.2582900.

Mansor, M. N., Jamil, S. H. F. S. M., Rejab, M. N. & Jamil, A. H. F. S. M. (2012). *K-nn algorithm for fast infant pain detection*. Conference Proceedings presented in 2012 International Symposium on Instrumentation & Measurement, Sensor Network and Automation (IMSNA), 25-28 Aug. 2012 (pp. 358-60). doi: 10.1109/MSNA.2012.6324593.

Mansor, M. N. B., Yaacob, S., Nagarajan, R. & Hariharan, M. (2010). *Detection of facial changes for hospital ICU patients using neural network*. Conference Proceedings presented in 2010 6th International Colloquium on Signal Processing & its Applications (pp. 1-4). doi: 10.1109/CSPA.2010.5545280.

Mansor, M. N. & Rejab, M. N. (2014). A Robust Neonatal Facial Pain Cues Classification. *Applied Mechanics and Materials*, 475-476, 1110-1117. doi: 10.4028/www.scientific.net/AMM.475-476.1110.

Martinez, M. & Stiefelhagen, R. (2013). *Automated Multi-Camera System for Long Term Behavioral Monitoring in Intensive Care Units*. Conference Proceedings presented in MVA (pp. 97-100).

Matsuda, N., Yamamoto, T., Miwa, M., Nukumi, S., Mori, K., Kuinose, Y., Maeda, E., Miura, H., Taki, H., Hori, S. & Abe, N. (2005). *Improvement of relief algorithm to prevent inpatient's downfall accident with night-vision CCD camera*. Conference Proceedings presented in Optomechatronic Machine Vision, December 5, 2005 - December 7, 2005 (pp. SPIE - The International Society for Optical Engineering; Hokkaido University, Japan; Sapporo International Plaza, Japan). doi: 10.1117/12.648639.

Matthews, G., Middleton, W., Gilmartin, B. & Bullimore, M. (1991). Pupillary diameter and cognitive load. *Journal of Psychophysiology*.

Meng, C. & Zhang, T. (2016). *Human Eye detection via sparse representation*. Conference Proceedings presented in 4th International Conference on Communications, Signal Processing, and Systems, CSPS 2015, October 23, 2015 - October 24, 2015 (pp. 693-700). doi: 10.1007/978-3-662-49831-6_71.

Meng, S. H., Hu, S. B., Huang, A. C., Huang, T. J., Xie, Z. & Jian, C. (2018). *Research on Eye Detection and Fatigue Early Warning Technologies*. Conference Proceedings presented in Proceedings of the Fourth Euro-China Conference on Intelligent Data Analysis and Applications (pp. 3-9).

Merck Manuals. (2017). Traumatic Brain Injury [Web Page]. Consulted at https://www.merckmanuals.com/professional/injuries-poisoning/traumatic-brain-injury -tbi/traumatic-brain-injury-tbi.

Merjulah, R. & Chandra, J. (2017). *Segmentation technique for medical image processing: A survey*. Conference Proceedings presented in 2017 International Conference on Inventive Computing and Informatics (ICICI) (pp. 1055-1061). doi: 10.1109/ICICI.2017.8365301.

Merkel, S. I., Voepel-Lewis, T., Shayevitz, J. R. & Malviya, S. (1997). The FLACC: a behavioral scale for scoring postoperative pain in young children. *Pediatric nursing*, 23(3), 293-297. Consulted at https://www.scopus.com/inward/record.uri?eid=2-s2.0 -0031134960&partnerID=40&md5=fd07b7c3128542adf0566daf03e0ea59.

Merkel, S., Voepel-Lewis, T. & Malviya, S. (2002). Pain Assessment in Infants and Young Children: The FLACC Scale: A behavioral tool to measure pain in young children. *AJN The American Journal of Nursing*, 102(10), 55-58.

Ming-Hsuan, Y., Kriegman, D. J. & Ahuja, N. (2002). Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 34-58. doi: 10.1109/34.982883.

Mingxin, Y., Yingzi, L. & Xiangzhou, W. (2016). An efficient hybrid eye detection method. *Turkish Journal of Electrical Engineering & Computer Sciences*, 24(3), 1586-603. doi: 10.3906/elk-1312-150.

Miyakawa, M. & Kosugi, T. (2003). *A three-dimensional gazing point detection system for communication analysis*. Conference Proceedings presented in Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439) (pp. 918-921 Vol.1). doi: 10.1109/IEMBS.2003.1279915.

Monwar, M., Rezaei, S. & Prkachin, K. (2007). Eigenimage based pain expression recognition. *IAENG International Journal of Applied Mathematics*, 36(2), 1-6.

Monwar, M. M. & Rezaei, S. (2006). *Pain recognition using artificial neural network*. Conference Proceedings presented in Signal Processing and Information Technology, 2006 IEEE International Symposium on (pp. 28-33).

Monzo, D., Albiol, A., Sastre, J. & Albiol, A. (2011). Precise eye localization using HOG descriptors. *Machine Vision and Applications*, 22(3), 471-480. doi: 10.1007/s00138-010-0273-0.

Morray, J. P., Tyler, D. C., Jones, T. K., Stuntz, J. T. & Lemire, R. J. (1984). Coma scale for use in brain-injured children. *Critical care medicine*, 12(12), 1018-1020.

Munn, S. M. (2009). *Three-dimensional head motion, point-of-regard and encoded gaze fixations in real scenes: Next-Generation portable video-based monocular eye tracking*. (Thesis).

Mutneja, V. & Singh, S. (2017). Modified Viola–Jones algorithm with GPU accelerated training and parallelized skin color filtering-based face detection. *Journal of Real-Time Image Processing*, 1-21.

Naji, S., Jalab, H. A. & Kareem, S. A. (2018). A survey on skin detection in colored images. *Artificial Intelligence Review*, 1-47.

Nanni, L., Brahnam, S. & Lumini, A. (2010a). A local approach based on a Local Binary Patterns variant texture descriptor for classifying pain states. *Expert Systems with Applications*, 37(12), 7888-7894. doi: 10.1016/j.eswa.2010.04.048.

Nanni, L., Lumini, A. & Brahnam, S. (2010b). Local binary patterns variants as texture descriptors for medical image analysis. *Artificial intelligence in medicine*, 49(2), 117-125. doi: 10.1016/j.artmed.2010.02.006.

Nejatian, A. & Sarbishei, G. (2017). *Implementation real-time gender recognition based on facial features using a hybrid neural network Imperialist Competitive Algorithm*. Conference Proceedings presented in 2017 Iranian Conference on Electrical Engineering (ICEE), 2-4 May 2017 (pp. 1584-9). doi: 10.1109/IranianCEE.2017.7985298.

Nelson, L. P., Lachman, S. E. & Gold, J. I. (2017). 2.33 Diagnosing Delirium in a Pediatric Intensive Care Unit. *Journal of the American Academy of Child & Adolescent Psychiatry*, 56(10, Supplement), S188-S189. doi: https://doi.org/10.1016/j.jaac.2017.09.113.

Newton, C., Chokwe, T., Schellenberg, J. A., Winstanley, P., Forster, D., Peshu, N., Kirkham, F. & Marsh, K. (1997). Coma scales for children with severe falciparum malaria. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 91(2), 161-165. doi: 10.1016/s0035-9203(97)90207-8.

Ngxande, M., Tapamo, J.-R. & Burke, M. (2019). Detecting inter-sectional accuracy differences in driver drowsiness detection algorithms. *arXiv preprint arXiv:1904.12631*.

Ni, B., Nguyen, C. D. & Moulin, P. (2012). *RGBD-camera based get-up event detection for hospital fall prevention*. Conference Proceedings presented in 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1405-1408).

Norimatsu, Y., Mita, S., Kozuka, K., Nakano, T. & Yamamoto, S. (2003). *Detection of the gaze direction using the time-varying image processing*. Conference Proceedings presented in Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems (pp. 74-79 vol.1). doi: 10.1109/ITSC.2003.1251924.

Oakes, L. M. (2012). Advances in Eye Tracking in Infancy Research. *Infancy*, 17(1), 1-8. doi: doi:10.1111/j.1532-7078.2011.00101.x.

Obinata, G., Tokuda, S., Fukuda, K. & Hamada, H. (2009). *Quantitative evaluation of mental workload by using model of involuntary eye movement*. Conference Proceedings presented in Engineering Psychology and Cognitive Ergonomics. 8th International Conference, (EPCE 2009). Held as Part of HCI International 2009, 19-24 July 2009 (pp. 223-32). doi: 10.1007/978-3-642-02728-4_24.

Ojo, J. & Adeniran, S. (2013). Illumination invariant face detection using hybrid skin segmentation method. *European Journal of Computer Science and Information Technology*, 1(4), 1-9.

Okada, Y. & Murashta, K. (2015). Development of a bed detection method for a camera-based patient monitoring system in hospitals and care facilities. *Seimitsu Kogaku Kaishi/Journal of the Japan Society for Precision Engineering*, 81(12), 1127-1132. doi: 10.2493/jjspe.81.1127.

Otten, P., Son, S. H. & Kim, J. (2014). *Automating Stroke Patient Evaluation Using Sensor Data and SVM*. Conference Proceedings presented in 2014 IEEE 7th International Conference on Service-Oriented Computing and Applications (pp. 223-229). doi: 10.1109/SOCA.2014.29.

Pai, Y.-T., Ruan, S.-J., Shie, M.-C. & Liu, Y.-C. (2006). *A simple and accurate color face detection algorithm in complex background*. Conference Proceedings presented in Multimedia and Expo, 2006 IEEE International Conference on (pp. 1545-1548).

Parodi, E., Melis, D., Boulard, L., Gavelli, M. & Baccaglini, E. (2017). *Automated newborn pain assessment framework using computer vision techniques*. Conference Proceedings presented in 2017 International Conference on Bioinformatics Research and Applications, ICBRA 2017, December 8, 2017 - December 10, 2017 (pp. 31-36).

doi: 10.1145/3175587.3175590.

Phillips, C. L., Bruno, M.-A., Maquet, P., Boly, M., Noirhomme, Q., Schnakers, C., Vanhaudenhuyse, A., Bonjean, M., Hustinx, R., Moonen, G., Luxen, A. & Laureys, S. (2011). "Relevance vector machine" consciousness classifier applied to cerebral metabolism of vegetative and locked-in patients. *NeuroImage*, 56(2), 797-808. doi: https://doi.org/10.1016/j.neuroimage.2010.05.083.

Phung, S. L., Bouzerdoum, A. & Chai, D. (2005). Skin segmentation using color pixel classification: analysis and comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1), 148-154. doi: 10.1109/TPAMI.2005.17.

Poppe, R. (2010). A survey on vision-based human action recognition. *Image and vision computing*, 28(6), 976-990. doi: 10.1016/j.imavis.2009.11.014.

Prati, A., Shan, C. & Wang, K. I.-K. (2019). Sensors, vision and networks: From video surveillance to activity recognition and health monitoring. *Journal of Ambient Intelligence and Smart Environments*, 11(1), 5-22. doi: 10.3233/Ais-180510.

Prochazka, A., Schatz, M., Vysata, O. & Valis, M. (2016). Microsoft Kinect Visual and Depth Sensors for Breathing and Heart Rate Analysis. *Sensors (Basel)*, 16(7), 996. doi: 10.3390/s16070996.

Prochazka, A., Charvatova, H., Vysata, O., Kopal, J. & Chambers, J. (2017). Breathing Analysis Using Thermal and Depth Imaging Camera Video Records. *Sensors (Basel)*, 17(6), 1408. doi: 10.3390/s17061408.

PyImageSearch. (2015). Local Binary Patterns with Python & OpenCV [Web Page]. Consulted at https://www.pyimagesearch.com/2015/12/07/local-binary-patterns-with-python-opencv/.

Qiong, L. & Guang-zheng, P. (2010). *A robust skin color based face detection algorithm*. Conference Proceedings presented in 2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010) (pp. 525-528). doi: 10.1109/CAR.2010.5456614.

Radwan, A. A., El-Bakry, H. M. & El Hadad, H. M. (2011). *A new expert system for pediatric respiratory diseases by using neural networks*. Conference Proceedings presented in 2nd International Conference on Applied Informatics and Computing Theory, AICT'11, September 26, 2011 - September 28, 2011 (pp. 296-306).

Ramsay, M., Savege, T., Simpson, B. & Goodwin, R. (1974). Controlled sedation with alphaxalone-alphadolone. *British medical journal*, 2(5920), 656.

doi: 10.1136/bmj.2.5920.656.

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016). *You only look once: Unified, real-time object detection*. Conference Proceedings presented in Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

Reilly, P. L., Simpson, D. A., Sprod, R. & Thomas, L. (1988). Assessing the conscious level in infants and young children: a paediatric version of the Glasgow Coma Scale. *Child's Nervous System*, 4(1), 30-33. doi: 10.1007/bf00274080.

Rein-Lien, H., Abdel-Mottaleb, M. & Jain, A. K. (2002). Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 696-706. doi: 10.1109/34.1000242.

Ren, S., He, K., Girshick, R. & Sun, J. (2015). *Faster r-cnn: Towards real-time object detection with region proposal networks*. Conference Proceedings presented in Advances in neural information processing systems (pp. 91-99).

Rezaei, M. & Klette, R. (2011). *3D Cascade of Classifiers for Open and Closed Eye Detection in Driver Distraction Monitoring*. Conference Proceedings presented in Computer Analysis of Images and Patterns (pp. 171-179).

Saeijs, R. W. J. J., Tjona Ten, W. E. & De With, P. H. N. (2018). *Dual-camera 3D head tracking for clinical infant monitoring*. Conference Proceedings presented in 2018 International Conference on Intelligent Systems and Computer Vision, ISCV 2018, April 2, 2018 - April 4, 2018 (pp. 1-8). doi: 10.1109/ISACV.2018.8354068.

Saeijs, R. W. J. J., Tjona Ten, W. E. & De With, P. H. N. (2019). *Dual-Camera Facial Landmark Tracking for Clinical Infant Monitoring*. Conference Proceedings presented in 14th International Conference on Signal Image Technology and Internet Based Systems, SITIS 2018, November 26, 2018 - November 29, 2018 (pp. 151-158). doi: 10.1109/SITIS.2018.00032.

Saez de Urabain, I. R., Nuthmann, A., Johnson, M. H. & Smith, T. J. (2017). Disentangling the mechanisms underlying infant fixation durations in scene perception: A computational account. *Vision Research*, 134, 43-59. doi: https://doi.org/10.1016/j.visres.2016.10.015.

Salvucci, D. D. & Goldberg, J. H. (2000). *Identifying fixations and saccades in eye-tracking protocols*. Conference Proceedings presented in Proceedings of the Eye Tracking Research and Applications Symposium 2000, November 6, 2000 - November 8, 2000 (pp. 71-78).

Sanders, R., Ma, D., Brooks, P. & Maze, M. (2008). Balancing paediatric anaesthesia: preclinical insights into analgesia, hypnosis, neuroprotection, and neurotoxicity. *British journal of anaesthesia*, 101(5), 597-609. doi: 10.1093/bja/aen263.

Satoh, H., Takeda, F., Saeki, Y., Ikeda, R. & Shiraishi, Y. (2006). *Proposal of awakening behavior detection system using neural network*. Conference Proceedings presented in 2nd IASTED International Conference on Computational Intelligence, CI 2006, November 20, 2006 - November 22, 2006 (pp. 164-169).

Savakis, A., Sharma, R. & Kumar, M. (2014). *Efficient eye detection using HOG-PCA descriptor*. Conference Proceedings presented in Imaging and Multimedia Analytics in a Web and Mobile World 2014 (pp. 90270J).

Schieveld, J. N., Leentjens, A. F. & Jellinck, M. S. (2005). Delirium in severely ill young children in the pediatric intensive care unit (PICU). *Journal of the American Academy of Child & Adolescent Psychiatry*, 44(4), 392-394. doi: 10.1097/01.chi.0000153231.64968.1a.

Scikit-learn. (n.d.). Support Vector Machines [Web Page]. Consulted at https://scikit-learn.org/stable/modules/svm.html.

Seshia, S., Johnston, B. & Kasian, G. (1983). Non-traumatic coma in childhood: clinical variables in prediction of outcome. *Developmental Medicine & Child Neurology*, 25(4), 493-501. doi: 10.1111/j.1469-8749.1983.tb13796.x.

Sessler, C. N., Gosnell, M. S., Grap, M. J., Brophy, G. M., O'neal, P. V., Keane, K. A., Tesoro, E. P. & Elswick, R. (2002). The Richmond Agitation–Sedation Scale: validity and reliability in adult intensive care unit patients. *American journal of respiratory and critical care medicine*, 166(10), 1338-1344. doi: 10.1164/rccm.2107138.

Shaik, K. B., Ganesan, P., Kalist, V., Sathish, B. & Jenitha, J. M. M. (2015). Comparative study of skin color detection and segmentation in HSV and YCbCr color space. *Procedia Computer Science*, 57, 41-48. doi: 10.1016/j.procs.2015.07.362.

Shim, J., Shim, M.-H., Baek, Y.-S. & Han, T.-D. (2011). *The development of a detection system for seniors' accidental fall from bed using cameras*. Conference Proceedings presented in Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication, ICUIMC 2011. doi: 10.1145/1968613.1968734.

Shishvan, O. R., Zois, D.-S. & Soyata, T. (2018). Machine intelligence in healthcare and medical cyber physical systems: A survey. *IEEE Access*, 6, 46419-46494. doi: 10.1109/Access.2018.2866049.

Shu, T., Zhang, B. & Tang, Y. Y. (2017). Novel Noninvasive Brain Disease Detection System Using a Facial Image Sensor. *Sensors (Basel)*, 17(12), 2843. doi: 10.3390/s17122843.

Sikka, K. (2014). *Facial Expression Analysis for Estimating Pain in Clinical Settings*. Conference Proceedings presented in Proceedings of the 16th International Conference on Multimodal

Interaction (pp. 349-353).

Sikka, K., Ahmed, A. A., Diaz, D., Goodwin, M. S., Craig, K. D., Bartlett, M. S. & Huang, J. S. (2015). Automated assessment of children's postoperative pain using computer vision. *Pediatrics*, 136(1), e124-e131. doi: 10.1542/peds.2015-0029.

Simpson, D., Cockington, R., Hanieh, A., Raftos, J. & Reilly, P. (1991). Head injuries in infants and young children: the value of the Paediatric Coma Scale. *Child's Nervous System*, 7(4), 183-190. doi: 10.1007/bf00249393.

Slanzi, G., Balazs, J. A. & VelÃ¡squez, J. D. (2017). Combining eye tracking, pupil dilation and EEG analysis for predicting web users click intention. *Information Fusion*, 35, 51-57. doi: https://doi.org/10.1016/j.inffus.2016.09.003.

Smith, H. A., Boyd, J., Fuchs, D. C., Melvin, K., Berry, P., Shintani, A., Eden, S. K., Terrell, M. K., Boswell, T. & Wolfram, K. (2011). Diagnosing delirium in critically ill children: Validity and reliability of the Pediatric Confusion Assessment Method for the Intensive Care Unit. *Critical care medicine*, 39(1), 150. doi: 10.1097/CCM.0b013e3181feb489.

Smith, H. A., Gangopadhyay, M., Goben, C. M., Jacobowski, N. L., Chestnut, M. H., Savage, S., Rutherford, M. T., Denton, D., Thompson, J. L., Chandrasekhar, R. et al. (2016). The Preschool Confusion Assessment Method for the ICU (psCAM-ICU): Valid and Reliable Delirium Monitoring for Critically Ill Infants and Children. *Critical care medicine*, 44(3), 592.

Sohail, A. S. M. & Bhattacharya, P. (2007). *Classification of Facial Expressions Using K-Nearest Neighbor Classifier*. Conference Proceedings presented in Computer Vision/Computer Graphics Collaboration Techniques (pp. 555-566).

Son Lam, P., Bouzerdoum, A. & Chai, D. (2002). *A novel skin color model in YCbCr color space and its application to human face detection*. Conference Proceedings presented in Proceedings. International Conference on Image Processing (pp. I-I). doi: 10.1109/ICIP.2002.1038016.

Song, F., Tan, X., Liu, X. & Chen, S. (2014). Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients. *Pattern Recognition*, 47(9), 2825-2838. doi: 10.1016/j.patcog.2014.03.024.

Soni, L. N., Datar, A. & Datar, S. (2017). Implementation of Viola-Jones Algorithm Based Approach for Human Face Detection. *International Journal of Current Engineering and Technology*, 7(5), 1819-1823.

Stern, J. A., Boyer, D. & Schroeder, D. (1994). Blink rate: a possible measure of fatigue. *Hum Factors*, 36(2), 285-97. doi: 10.1177/001872089403600209.

Story, T. N. (2016). *A comparison of fixation and fractal measures of eye movement when viewing pictures with affective valence*. (Thesis).

Strupczewski, A., Czuprynski, B., Naruniec, J. & Mucha, K. (2016). *Geometric Eye Gaze Tracking*. Conference Proceedings presented in VISIGRAPP (3: VISAPP) (pp. 446-457).

Sun, Y., Shan, C., Tan, T., Long, X., Pourtaherian, A. & Zinger, S. (2018). Video-based discomfort detection for infants. *Machine Vision and Applications*, 1-12.

Suprenant, S. & Coghlan, M. A. (2016). Respiratory Distress in the Newborn: An Approach for the Emergency Care Provider. *Clinical Pediatric Emergency Medicine*, 17(2), 113-121. doi: 10.1016/j.cpem.2016.03.004.

Sury, M. R. J. & Bould, M. D. (2011). Defining awakening from anesthesia in infants: a narrative review of published descriptions and scales of behavior. *Pediatric Anesthesia*, 21(4), 364-372. doi: doi:10.1111/j.1460-9592.2011.03538.x.

Sury, M. R. J. (2010). *Characterisation of awakening from anaesthesia in infants*. (Thesis).

Takeda, F. (2013). *Proposal of an awakening behavior detection system for medical use and adaptation for fluctuation of the brightness quantity with infrared camera device kinect*. Conference Proceedings presented in 2013 9th International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2013, December 2, 2013 - December 5, 2013 (pp. 714-719). doi: 10.1109/SITIS.2013.116.

Tan, W. R., Chan, C. S., Yogarajah, P. & Condell, J. (2012). A fusion approach for efficient human skin detection. *IEEE Transactions on Industrial Informatics*, 8(1), 138-147. doi: 10.1109/Tii.2011.2172451.

Tatman, A., Warren, A., Williams, A., Powell, J. E. & Whitehouse, W. (1997). Development of a modified paediatric coma scale in intensive care clinical practice. *Archives of Disease in Childhood*, 77(6), 519-521. doi: 10.1136/adc.77.6.519.

Teasdale, G. & Jennett, B. (1974). Assessment of coma and impaired consciousness: a practical scale. *The Lancet*, 304(7872), 81-84. doi: 10.1016/s0140-6736(74)91639-0.

Thompson, B. T., Chambers, R. C. & Liu, K. D. (2017). Acute respiratory distress syndrome. *New England Journal of Medicine*, 377(6), 562-572. doi: 10.1056/NEJMra1608077.

Tiawongsombat, P. & Rattanapoka, C. (2015a). A study of two robust features for effective open or closed eye classification. *Applied Mechanics and Materials*, 781, 507-10. doi: 10.4028/www.scientific.net/AMM.781.507.

Tiawongsombat, P. & Rattanapoka, C. (2015b). A Study of Two Robust Features for Effective Open or Closed Eye Classification. *Applied Mechanics and Materials*, 781, 507-510. doi: 10.4028/www.scientific.net/AMM.781.507.

Tiemeng, L., Wenjun, H., Fei, L., Yu, L. & Chen, X. (2016). *Face Detection Based on Depth Information Using HOG-LBP*. Conference Proceedings presented in 2016 Sixth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC), 21-23 July 2016 (pp. 779-84). doi: 10.1109/IMCCC.2016.92.

Tokuda, S., Obinata, G., Palmer, E. & Chaparro, A. (2011). *Estimation of mental workload using saccadic eye movements in a free-viewing task*. Conference Proceedings presented in 2011 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 30 Aug.-3 Sept. 2011 (pp. 4523-9). doi: 10.1109/IEMBS.2011.6091121.

Tong, X., Qin, H. & Zhuo, L. (2017). *An eye state recognition algorithm based on feature level fusion*. Conference Proceedings presented in 2017 IEEE International Conference on Vehicular Electronics and Safety (ICVES) (pp. 151-155). doi: 10.1109/ICVES.2017.7991917.

Toyama, K., Krumm, J., Brumitt, B. & Meyers, B. (1999). *Wallflower: Principles and practice of background maintenance*. Conference Proceedings presented in Proceedings of the Seventh IEEE International Conference on Computer Vision (pp. 255-261).

Traube, C., Silver, G., Kearney, J., Patel, A., Atkinson, T. M., Yoon, M. J., Halpert, S., Augenstein, J., Sickles, L. E. & Li, C. (2014). Cornell Assessment of Pediatric Delirium: a valid, rapid, observational tool for screening delirium in the PICU. *Critical care medicine*, 42(3), 656. doi: 10.1097/CCM.0b013e3182a66b76.

Tryolabs. (2018). Luminoth. Consulted at https://luminoth.ai/.

Urban, B. & Bleckwenn, M. (2002). Concepts and correlations relevant to general anaesthesia. *British journal of anaesthesia*, 89(1), 3-16. doi: 10.1093/bja/aef164.

Van Der Meer, E., Beyer, R., Horn, J., Foth, M., Bornemann, B., Ries, J., Kramer, J., Warmuth, E., Heekeren, H. R. & Wartenburger, I. (2010). Resource allocation and fluid intelligence: Insights from pupillometry. *Psychophysiology*, 47(1), 158-169. doi: 10.1111/j.1469-8986.2009.00884.x.

Van Dijk, M., Peters, J. W., Van Deventer, P. & Tibboel, D. (2005). The COMFORT Behavior Scale: a tool for assessing pain and sedation in infants. *AJN The American Journal of*

Nursing, 105(1), 33-36. doi: 10.1097/00000446-200501000-00019.

Vezhnevets, V., Sazonov, V. & Andreeva, A. (2003). *A survey on pixel-based skin color detection techniques*. Conference Proceedings presented in Proc. Graphicon (pp. 85-92).

Vieira, L. C. & Da Silva, F. S. C. (2016). *Assessment of fun from the analysis of facial expressions to support video game design*. Conference Proceedings presented in AISB Annual Convention 2016, AISB 2016. Consulted at https://www.scopus.com/inward/record.uri?eid=2-s2.0-85041896335&partnerID=40&md5=91d216d72ec865a1dd1696557b0b0fc2.

Viola, P. & Jones, M. (2001). *Rapid object detection using a boosted cascade of simple features*. Conference Proceedings presented in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001 (pp. I-511-I-518 vol.1). doi: 10.1109/CVPR.2001.990517.

Viola, P. & Jones, M. J. (2004). Robust Real-Time Face Detection. *International Journal of Computer Vision*, 57(2), 137-154. doi: 10.1023/B:VISI.0000013087.49260.fb.

Wang, C.-W., Hunter, A., Gravill, N. & Matusiewicz, S. (2014). Unconstrained video monitoring of breathing behavior and application to diagnosis of sleep apnea. *IEEE Transactions on Biomedical Engineering*, 61(2), 396-404. doi: 10.1109/TBME.2013.2280132.

Wass, S. V., Smith, T. J. & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods*, 45(1), 229-250. doi: 10.3758/s13428-012-0245-6.

Wei, Z. & Li-min, X. (2011). Pain expression recognition based on SLPP and MKSVM. *International Journal of Engineering and Manufacturing*, 3, 69-74.

Wenhui, D. & Peishu, Q. (2009). *Eye state classification based on multi-feature fusion*. Conference Proceedings presented in 2009 Chinese Control and Decision Conference (CCDC 2009), 17-19 June 2009 (pp. 231-4). doi: 10.1109/CCDC.2009.5195119.

Whinnery, T. & Forster, E. M. (2017). The first sign of loss of consciousness. *Physiology and Behavior*, 179, 494-503. doi: 10.1016/j.physbeh.2017.06.022.

Wilson, P. I. & Fernandez, J. (2006). Facial feature detection using Haar classifiers. *Journal of Computing Sciences in Colleges*, 21(4), 127-133.

Wu, J. & Zhou, Z.-H. (2003). Efficient face candidates selector for face detection. *Pattern recognition*, 36(5), 1175-1186. doi: Pii S0031-3203(02)00165-6 Doi 10.1016/S0031-3203(02)00165-6.

Wu, Y. & Ji, Q. (2015). Shape augmented regression method for face alignment. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 26-32. doi: 10.1109/Iccvw.2015.129.

Xie, Y., Chen, K. & Murphey, Y. L. (2018). *Real-time and Robust Driver Yawning Detection with Deep Neural Networks*. Conference Proceedings presented in 2018 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 532-538).

Xuan, Z., Xin, G., Jiajun, W., Hui, Y., Zhiyong, W. & Zheru, C. (2017). Eye tracking data guided feature selection for image classification. *Pattern Recognition*, 63, 56-70. doi: 10.1016/j.patcog.2016.09.007.

Yager, J. Y., Johnston, B. & Seshia, S. S. (1990a). Coma scales in pediatric practice. *Am J Dis Child*, 144(10), 1088-1091. doi: 10.1001/archpedi.1990.02150340032019.

Yager, J. Y., Johnston, B. & Seshia, S. S. (1990b). Coma scales in pediatric practice. *Am J Dis Child*, 144(10), 1088-1091. doi: 10.1001/archpedi.1990.02150340032019.

Yasova Barbeau, D. & Weiss, M. D. (2017). Sleep Disturbances in Newborns. *Children*, 4(10), 90. doi: 10.3390/children4100090.

Yimyam, W. & Ketcham, M. (2018). *Eye Region Detection in Fatigue Monitoring for the Military Using AdaBoost Algorithm*. Conference Proceedings presented in Advances in Natural Language Processing, Intelligent Informatics and Smart Technology (pp. 151-161).

Yis, U., Yüksel, Ç. N., Köse, A. & Erdogan, S. (2012). Evaluation of cases with impaired state of consciousness: Gaziantep Children's Hospital experience. *The Turkish journal of pediatrics*, 54(1), 30. Consulted at https://www.ncbi.nlm.nih.gov/pubmed/22397039.

Yu, X., Han, W., Li, L., Shi, J. Y. & Wang, G. (2011). *An eye detection and localization system for natural human and robot interaction without face detection*. Conference Proceedings presented in Conference Towards Autonomous Robotic Systems (pp. 54-65).

Zafeiriou, S., Zhang, C. & Zhang, Z. (2015). A survey on face detection in the wild: Past, present and future. *Computer Vision and Image Understanding*, 138, 1-24. doi: https://doi.org/10.1016/j.cviu.2015.03.015.

Zamzami, G., Ruiz, G., Goldgof, D., Kasturi, R., Yu, S. & Ashmeade, T. (2015). *Pain assessment in infants: Towards spotting pain expression based on infants' facial strain*. Conference Proceedings presented in 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) (pp. 1-5). doi: 10.1109/FG.2015.7284857.

Zamzmi, G., Pai, C.-Y., Goldgof, D., Kasturi, R., Sun, Y. & Ashmeade, T. (2016). Machine-based multimodal pain assessment tool for infants: a review. *arXiv preprint arXiv:1607.00331*.

Zangana, H. M. & Al-Shaikhli, I. F. (2013). A new algorithm for human face detection using skin color tone. *IOSR Journal of Computer Engineering*, 11(6), 31-38.

Zarit, B. D., Super, B. J. & Quek, F. K. (1999). *Comparison of five color models in skin pixel classification*. Conference Proceedings presented in Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999. Proceedings. International Workshop on (pp. 58-63).

Zhang, B., Vijaya kumar, B. V. & Zhang, D. (2014). Noninvasive diabetes mellitus detection using facial block color with a sparse representation classifier. *IEEE Trans Biomed Eng*, 61(4), 1027-33. doi: 10.1109/tbme.2013.2292936.

Zhang, C. & Zhang, Z. (2014). *Improving multiview face detection with multi-task deep convolutional neural networks*. Conference Proceedings presented in IEEE Winter Conference on Applications of Computer Vision (pp. 1036-1041). doi: 10.1109/WACV.2014.6835990.

Zhang, X., Gonnot, T. & Saniie, J. (2017). Real-Time Face Detection and Recognition in Complex Background. *Journal of Signal and Information Processing*, 08(02), 99-112. doi: 10.4236/jsip.2017.82007.

Zhang, X., Zhang, S. & Hapeshi, K. (2010). A new method for face detection in colour images for emotional bio-robots. *Science China Technological Sciences*, 53(11), 2983-2988. doi: 10.1007/s11431-010-4132-z.

Zhang, Y. & Ling, Q. (2017). Bicycle Detection Based On Multi-feature and Multi-frame Fusion in low-resolution traffic videos. *arXiv preprint arXiv:1706.03309*.

Zhao, L., Wang, Z., Wang, X. & Liu, Q. (2018). Driver drowsiness detection using facial dynamic fusion information and a DBN. *IET Intelligent Transport Systems*, 12(2), 127-133. doi: 10.1049/iet-its.2017.0183.

Zheng-Han, W. & Horng-Horng, L. (2017). *Eye detection using gradient histogram matching for cornea localization in refractive eye surgery*. Conference Proceedings presented in 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 14-16 Oct. 2017 (pp. 6 pp.). doi: 10.1109/CISP-BMEI.2017.8302191.