DEEP LEARNING SOLUTION FOR CHILDREN LONG-TERM IDENTIFICATION

*Nikolajs Bumanis[1], **Gatis Vitols[2]**, **Irina Arhipova[2]**, **Inga Meirane[2]**
[1]Latvia University of Life Sciences and Technologies, Latvia
[2]WeAreDots Ltd., Latvia
*Corresponding author's email: nikolajs.bumanis@llu.lv

Abstract
Deep learning algorithms are becoming default solution for application in business processes where recognition, identification and automated learning are involved. For human identification, analysis of various features can be applied. Face feature analysis is most popular method for identification of person in various stages of life, including children and infants.
The aim of this research was to propose deep learning solution for long-term identification of children in educational institutions. Previously proposed conceptual model for long-term re-identification was enhanced. The enhancements include processing of unexpected persons' scenarios, knowledge base improvements based on results of supervised and unsupervised learning, implementation of video surveillance zones within educational institutions and object tracking results' data chaining between multiple logical processes. Object tracking results are the solution we found for long-term identification realization.
Key words: neural netwoks, face recognition solutions, long-term person identification.

Introduction

Identifying or re-identifying a person is an essential aspect of identifying a person's identity and location by retrieving data from multiple devices (Arhipova, Vitols, & Meirane, 2020). Typical scenario in person identification is person descriptors (face characteristics, gait, etc.) extraction and comparison with previously created gallery.

Typically CCTV cameras are used to capture a person, and recognition must be performed within certain constraints, for example, different camera parameters, light variations, people specifics (facial expression, presence of facial hair, make up, aging, etc.).

If two images are taken with a few minutes or hours apart, one can assume that the person's visual appearance, for example, clothing will be the same. Such a re-identification scenario is called short-period Re-ID. If images or videos are taken at intervals of several days or months, the person's re-identification process is a long-period Re-ID (Bedagkar-Gala & Shah, 2014).

Deep learning algorithms are widely applied to a short term and long term person identification.

The aim of this research is to propose deep learning solution for long-term identification of children in schools.

Materials and Methods

Some researchers already were addressing person identification issues in controlled and non-controlled environment. For example, the solution proposed by Rothoft et al. (Rothoft *et al.*, 2017) allows to monitor children behaviour in class. To detect if children pay attention to the designed spatial area, distribution of the focus points in two dimensions are performed. After that identification of anomalous points is identified.

From deep learning algorithms, Convolutional neural network (CNN) and its derivatives, seems to be most applied as a core component for short term and long term re-identification solutions. For example, CNN network was compared with more classical and widely applied linear discriminant analysis (LDA) and support vector machines (SVM) for recognition task of 138 male and 112 female school age children data set (Vařeka, 2020). For such a small data set, all three algorithms showed similar performance of 62–64% for a single trial. However, it is known that CNN is more effective with larger data sets (Chauhan & Singh, 2019).

Researchers also address identification of abnormal activities of humans in crowded places using deep learning. For example, researchers propose
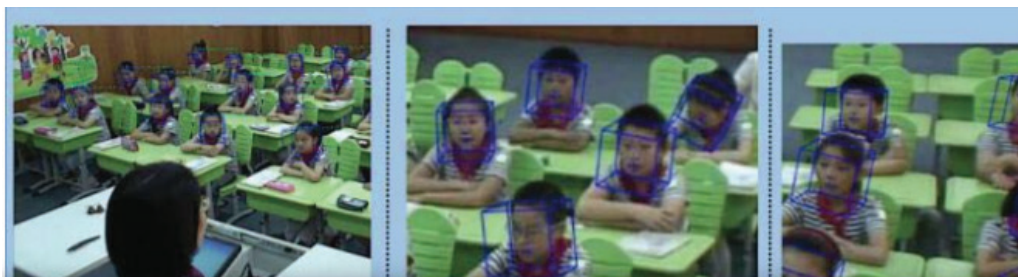

Figure 1. Focus points corresponding to inattentive students (Rothoft *et al*., 2017).
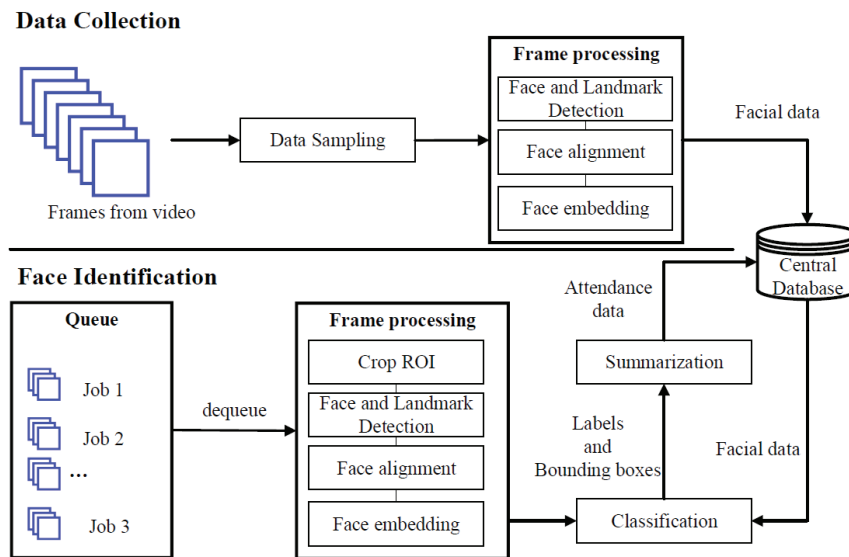
Figure 2. Components of face recognition module (Son *et al.*, 2020b).

to introduce robots (instead of classic CCTV) to identify abnormal activities in crowded places using CNN network and SVM for image classification task (Du, 2020). For training data UMN and PETS 2009 data set is used. Proposal showed improved accuracy and promising future implementations.

Gait is also an important feature for identification purposes and it is less affected by clothes, sunglasses, masks or other constraints that can impact recognition accuracy. Authors propose gait recognition method with CNN model and apply it using CASIA and OpenCV data sets. Results are applied to identify laboratory attendance with possible applications, for example, to record school attendance (Xu, Ren, & Nishide, 2020). Downside of gait recognition is that it is difficult to acquire data for recognition – capture the gait of the person.

For children identification, mostly face recognition is used. Face recognition has been applied to find missing children, support school safety and even social network activities of children (Oo & Oo, 2019). Children face long-term recognition is a challenging task, because children face is hard to recognize. Authors use three CNN for children face recognition to evaluate accuracy of networks, one VGG Face based and two VGG16 and ResNet50 based. Research authors conclude that MobileFaceNet data set provide best examples to train algorithms with improved accuracy (Oo & Oo, 2019). Children attendance to examinations to identify authenticity of person who takes the test has been also researched with implementation of deep learning functionality into schools CCTV systems (Othman & Aydin, 2019).

Some authors propose solutions that can be integrated with schools CCTV systems to record attendance of lectures or school in general (Son *et al.*, 2020).

Figure 2 illustrates proposed components of face recognition module. Face identification takes a photo from a queue and passes it to identification. Frame processing crops only part of image, providing identification of face in the image. Authors used face-embedding technique for feature description. All crops are passed to vectorization and then to classification tasks. After that, a summarization algorithm is applied which allows making the final decision. Data collection part in Figure 2 is responsive to the frame extraction and then frame processing is done building feature vectors and storing them in the central database. To acquire training data set for this solution, authors built mobile application to record face video maximum 30 seconds long. Videos are uploaded to database server for further processing via web service.

In the next section, we propose children long term re-identification model that incorporates some of components mentioned in this section.

Results and Discussion

Long term identification conceptual model (Figure 3) was proposed in our first paper (Arhipova, Vitols, & Meirane, 2020b) related to children face identification. The conceptual model was divided into four steps – video feed processing, creation of biometric pattern and person's identification, supervised and unsupervised learning. Video feed processing uses CNN network for frame identification purposes.

During the recent development of Re-ID prototype, the proposed conceptual model was critically analyzed from the perspective of compatibility degree versus
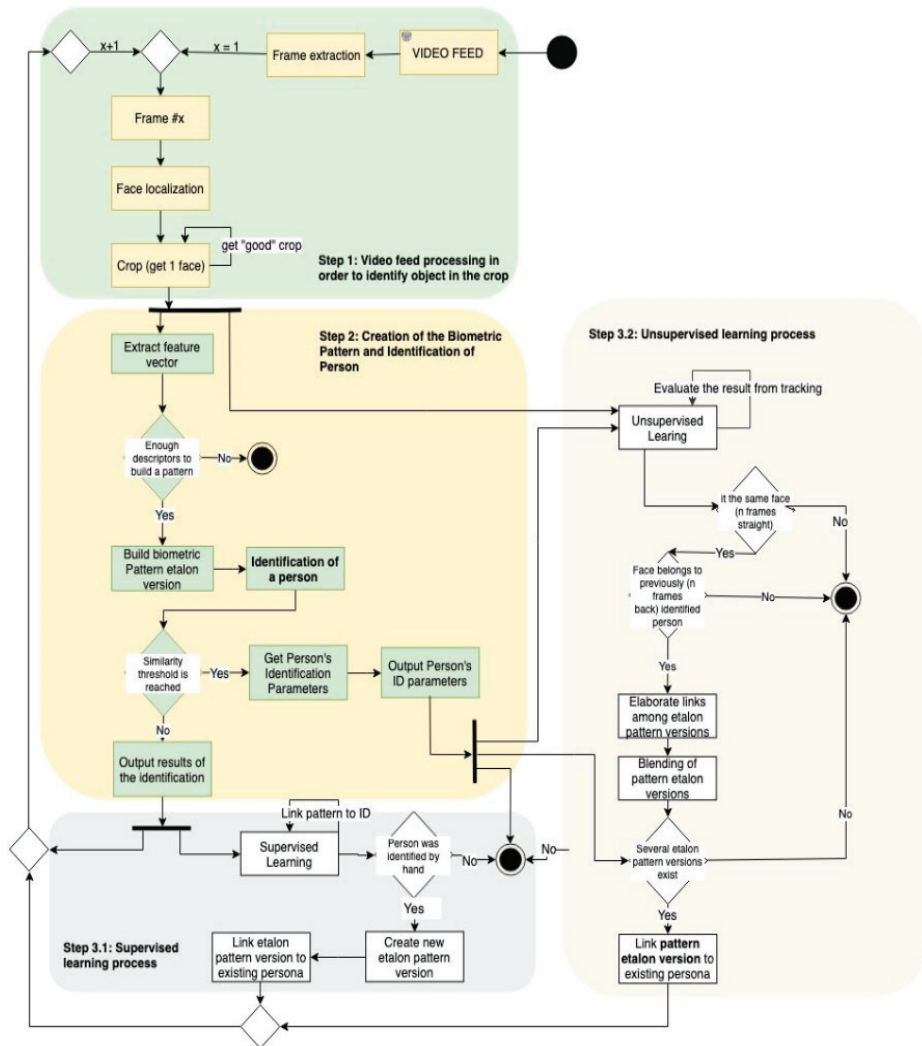
Figure 3. Re-ID conceptual model (Arhipova *et al*., 2020b).

real life application. To start with, this conceptual model does not include any mentioning of knowledge base creation prior to identification process execution, as well as potential management functionality. The knowledge base must be created using children biometric and biographic data. In order to log the results of object detection and person identification, a registration process must be included. The process of logging object detection must be put at the end of the step 1, whereas person identification logging must be executed at the end of the step 2 regardless of identification success. Both of these logs must contain reference data, i.e. the person identified must also be detected before that. The prototype for children attendance implies automatic fixation of children attendance during a lesson; the model, however, does not contain such a process, and, therefore, must be improved. We suggest adding this process after the previously mentioned person identification logging process at the end of the step 2. Finally, any management processes, including creation

and management of user accounts, management of existing biometric and biographic data, are out of this model; however, during a development of real life application the development of mentioned processes must be elaborated.

Conceptual model improvement

There are four aspects of real life application of long-term identification important for successful implementation. These aspects refer to improvements of functionality and scalability of Re-ID system, and include object tracking, processing the cases of unexpected person, integration with external systems, implementation of video zones in educational institutions.

Video surveillance area in an educational institution

The space within educational institution is divided into several video surveillance zones (hereinafter – zones) – a particular classroom, a specific hallway, etc. that are surveilled, as a stream, by one or more video cameras. It is assumed that each stream is handled by separate logical model process, and it

should be possible to initiate multiple streams from a single workstation, i.e. corridor surveillance from a security workstation.

Zone is defined by (1) restricted space, i.e. classroom, hallway or adjacent grounds belonging to institution, (2) video surveillance devices, and (3) workstation handling the video surveillance stream, provided that one workstation is capable of handling multiple streams simultaneously.

To implement a new zone within institution the sufficient number of video surveillance devices must be installed and a workstation must be assigned to handle the video surveillance stream, produced by these devices.

Object tracking

The proposed conceptual model was intending the use of object tracking in order to provide consistency and accuracy for this object's identification. In the framework of Re-ID system, the object is a person that can be educational institution's student, employee or visitor. The object tracking is performed to fulfill two objectives: (1) to conclude that the person within particular image's coordinates is the same and (2) to conclude that the person detected has changed location, but the system can still tell that it is the same person. Achieving the first objective provides information for precise logging of object detection period, while the second gives possibilities to track person across all video-monitoring zones, i.e. tell if a student attended the correct class, or the same student was doing something suspicious outdoors on educational institution's grounds. The realization of the second objective is also one of the long-term identification's goals.

In addition, object tracking allows tracking the motion of a person, which may be important to assess the quality of education. Motion detection can process the motion of children's heads, arms and body language in order to determine if they give sufficient attention to teachers during classes, to detect any aggressive behavior and to detect any noticeable health conditions.

Identification of an unexpected person

To provide a secure learning environment, any unexpected person must be detected. There are risks concerning an unexpected person. An unexpected person that is attending educational institution may not be this institution's student, employee or attending parent/guardian, and is attending it with malicious intents, i.e. stealing, bullying and selling prohibited stuff. An unexpected person, that is this institution's student, is attending another group's class, thus potentially worsening learning process for other students.

We developed solutions to address these risks. Firstly, assuming that all educational institution's students and employees are registered in the knowledge base, any person that cannot be identified for a period is considered an unexpected person. Secondly, any student that is attending other group's class is considered an unexpected person. Thirdly, information about each detected unexpected person is sent to school's security personnel. Lastly, information about an unexpected person, which is not institution's student, employee or attending parent/guardian and which is attending educational institution without justification repeatedly, is sent to relevant authorities.

Improvements to the conceptual model consist of adding a new process that is responsible for delivering
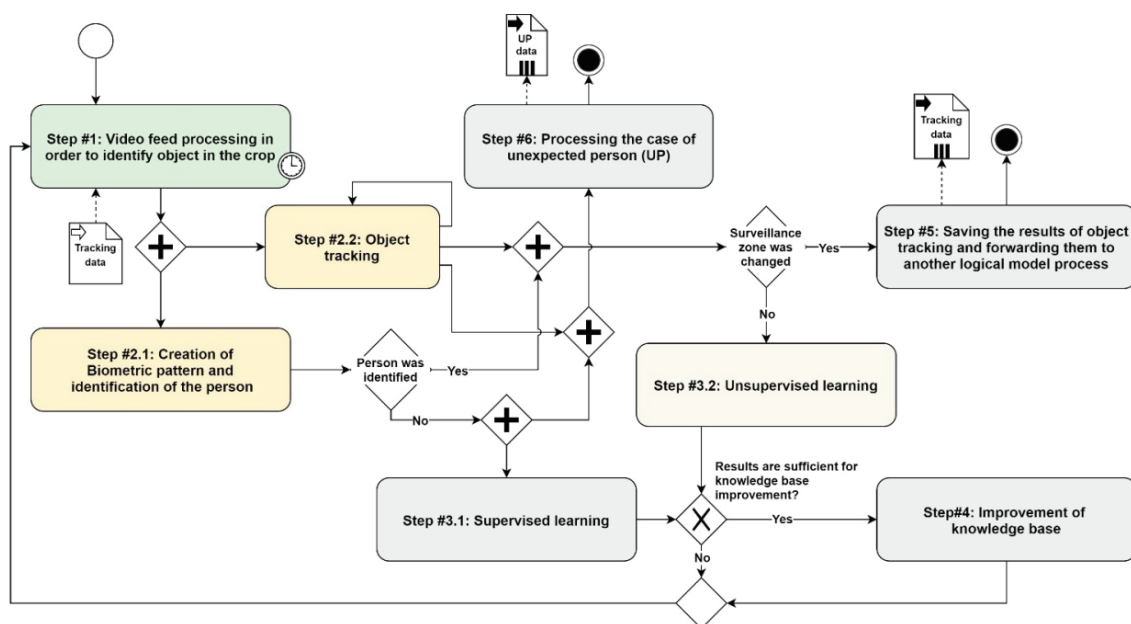

Figure 4. Improvements of conceptual Re-ID model.

information about a detected unexpected person to institution's security internal system.

Integration with external systems

The external systems are integrated into Re-ID system using input and output service gateways. We identified three systems related to educational institution: (1) educational institution's internal system, (2) educational institution management information system and (3) city's municipality information system.

Input data received from external systems include such data as (1) student's name, surname, personal number, birthdate, gender and photo, (2) student's social status, (3) information about student's parents/guardian. Student's social status can be used to classify students into groups and tailor educational process to particular group's needs. Input data about students received from the city's municipal system can be used to build a knowledge base, provided that this data is sufficient to establish the student's biographical and biometric identities.

Output data is generated automatically in order to reduce the needs for manual input, resulting in more efficient work for educational institution's employees. In the long term, the results of the identification give insight into whether the students are safe during classes or outside the educational institution.

We identified several output data, assuming that all students and employees are registered in the Re-ID knowledge base, and video-monitoring covers institution's classrooms, corridors and adjacent areas. These data include (1) attendance results that are logged in the institution's internal system's class attendance log, (2) the identities and total number of students who did not attend the institution on a daily/weekly/monthly basis, used for statistical analysis by institution's management information system, and (3) the biography data of students who have been identified doing suspicious activities, i.e. smoking, consuming alcohol, sent to institution's security system.

Model performance improvements and long-term re-identification

According to the scenarios described above, we developed a set of improvements (Figure 4) to the performance of the model.

These improvements are meant to address an object, respectively – a person, tracking across multiple surveillance zones, provided that a person is inbounds of overall surveillance coverage, enhancement of knowledge base using the results of supervised/unsupervised learning and realization of long-term re-identification.

Object tracking process is a repeatable process, executed after the person is first detected in a video stream. This process continues until a detected person cannot be repeatedly detected within the original surveillance zone, i.e. a person leaves a classroom. In case the process was terminated due to the change of surveillance zone and the detected person was successfully identified, the tracking data and information about the identified person are provided to another logical model process, which in turn uses these data during the next processing iteration. If it was impossible to identify the detected person, the information about this person is processed according to unexpected person's processing normative/methods.

Knowledge base is enhanced using the results of supervised and unsupervised learning. These results include additional feature value for existing feature vectors, as well as completely new feature vectors. The execution of these processes is determined by the surveillance zone, i.e. there is no need to perform supervised learning for a sport hall, whereas unsupervised learning within it may result in ambiguous and imprecise data. The fact of person's identification successfulness is determinant for choosing correct learning approach. Supervised learning is executed in case a person was not identified, and includes manual input from an expert, whereas unsupervised learning is performed for the identified person, provided that the surveillance zone was not changed for a particular period. The change of surveillance zone leads to the change of video environment – background texture, brightness, light angle, etc., and, therefore, can hardly produce correlated data.

The aspects of object tracking and knowledge base enhancement within the process of identification play the major role for the long-term re-identification. Object tracking data allows linking multiple logical processes (streams) to track object in space and time, whereas enhancement of the knowledge base enables the storage of several versions of the biometric identity etalon versions, which allows the person to be identified even after a relatively long period of time and in another surveillance area.

Conclusions

1. From deep learning algorithms, Convolutional neural network (CNN) and its derivatives seem to be most applied as a core component for short term and long term re-identification solutions.
2. For children identification, mostly face recognition methods are applied which are usually integrated with CCTV systems or use data collected from CCTV cameras.
3. There are usually two options to train the networks – with existing data or data published online. From available data sets MobileFaceNet show promising results in training accuracy.
4. Previously proposed long-term re-identification model was incomplete and was limited in terms

of scalability. Enhancement in the aspects of unexpected person processing and surveillance zone management propels the model in terms of both scalability and functionality.

5. Enhanced model can be used as a basis for development of health surveillance system. This can be accomplished by using infrared cameras and CNN, trained accordingly.

References

Arhipova, I., Vitols, G., & Meirane, I. (2020). *Long Period Re-identification Approach to Improving the Quality of Education: A Preliminary Study*. DOI: 10.1007/978-3-030-39442-4_14.

Bedagkar-Gala, A., & Shah, S.K. (2014). A survey of approaches and trends in person re-identification. *Image and Vision Computing*, 32(4), 270–286. DOI: 10.1016/j.imavis.2014.02.001.

Chauhan, N.K., & Singh, K. (2019). A review on conventional machine learning vs deep learning. *2018 International Conference on Computing, Power and Communication Technologies, GUCON 2018*, 347–352. DOI: 10.1109/GUCON.2018.8675097.

Du, Y. (2020). An anomaly detection method using deep convolution neural network for vision image of robot. *Multimedia Tools and Applications*. DOI: 10.1007/s11042-020-08684-1.

Oo, S.L.M., & Oo, A.N. (2019). Child Face Recognition with Deep Learning. *2019 International Conference on Advanced Information Technologies, ICAIT 2019*, 155–160. DOI: 10.1109/AITC.2019.8921152.

Othman, N.A., & Aydin, I. (2019). A Smart School by Using an Embedded Deep Learning Approach for Preventing Fake Attendance. *2019 International Conference on Artificial Intelligence and Data Processing Symposium, IDAP 2019*. DOI: 10.1109/IDAP.2019.8875883.

Rothoft, V., Si, J., Jiang, F., & Shen, R. (2017). Monitor Pupils' Attention by Image Super-Resolution and Anomaly Detection. *2017 International Conference on Computer Systems, Electronics and Control (ICCSEC)*, 843–847. Retrieved March 17, 2020, from https://ieeexplore.ieee.org/document/8446759/.

Son, N.T., Anh, B.N., Ban, T.Q., Chi, L.P., Chien, B.D., Hoa, D.X., … Khan, M.H.R. (2020). Implementing CCTV-based attendance taking support system using deep face recognition: A case study at FPT polytechnic college. *Symmetry*, 12(2). DOI: 10.3390/sym12020307.

Vařeka, L. (2020). Evaluation of convolutional neural networks using a large multi-subject P300 dataset. *Biomedical Signal Processing and Control*, 58. DOI: 10.1016/j.bspc.2019.101837.

Xu, Y., Ren, F., & Nishide, S. (2020). Using CNN's gait recognition to strengthen laboratory safety supervision. *Proceedings of 2019 the 9th International Workshop on Computer Science and Engineering, WCSE 2019*, 181–188. Hong-Kong.