Creating Value from Deep Learning Technology and Its Business Applications

Kunihiro Moriga, Takeharu Eda, Masashi Toyama, Keita Mikami, Yutaka Hirokawa, Yuji Yamada, Sanae Muramatsu, Taku Sasaki, Shin'ya Yamaguchi, and Katsuo Inaya

Abstract

NTT has positioned the utilization of artificial intelligence (AI) as an important business strategy. NTT Software Innovation Center has been engaged in research and development aimed at creating value from the AI technology known as *deep learning*. In this article, we explain how the deep-learning inference environment can be optimized for business applications, namely by speeding it up, making processing lightweight, and saving labor. We also report on verification activities concerning the image-analysis business opportunities (such as surveillance-camera image analysis) generated by using that technology.

Keywords: deep learning, cloud-based inference, optimization

1. Deep learning: entering the disillusionment phase

Artificial intelligence (AI) is currently said to be in its third boom. The Hype Cycle for Emerging Technologies, 2018 [1] published by Gartner Inc. in August 2018, predicted that deep learning would be at a *peak of inflated expectation* in its second consecutive year. In other words, it would soon enter a *disillusionment phase*.

Conditions such as PoC (proof of concept) and precedent cases and best practices were published by cutting-edge companies; however, the persons and departments in charge might be feeling the difficulty of solving problems using deep learning technology by themselves.

During the peak of inflated expectations phase, the results and utility that everyone imagined and expected cannot be obtained, and the people who are in charge become disappointed as they face the real situation; this is the beginning of the disillusionment phase. However, that situation can be said to be the beginning of true business applications. From now onwards, implementation and peripheral technologies of deep learning will catch up, and deep learning will gradually be adopted in actual business operations.

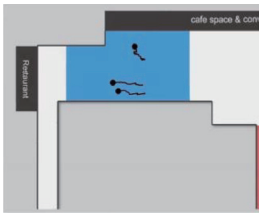
The technology developed by NTT Software Innovation Center that can analyze images of people in real time is an example of such technology that is approaching the stage of business application.

2. Analysis of images captured by multiple surveillance cameras at high speed in real time

Since deep learning became a topic of research in about 2011, it has succeeded in giving something akin to a *person's eyes and ears* to computers. Moreover, as of 2018, it is no exaggeration to say that at the purely technical level, it has already passed beyond the human eyes and ears ability. Real-time person tracking developed by NTT Software Innovation

When people are detected in the video image, they are enclosed in a blue frame. The line extending from the bottom of the blue frame is the estimated trajectory.

(a) People detection



As shown in the figure above, if a plan view of the room is prepared in advance, it is possible to map the trajectory onto the plan view.

(c) Trajectory recognition

After people are detected in the video image by the people detection function, if a photo (as shown on the right side of the figure) of a person being looked for is specified, when the specified person is detected in the image, the blue frame switches to a red frame, which indicates that the specified person has been found. Moreover, in this video, it is possible to identify the same person even with a backward-looking video by performing the whole-body collation function described in section 2.1.

(b) Person re-identification

Fig. 1. Functions of real-time person tracking.

Center is a form of deep learning packaged as image-analysis technology [2]. It analyzes images captured by a large number of surveillance cameras installed in a facility in real time and instantly detects and tracks target persons (suspicious persons, prominent persons, people needing care, lost children, etc.) in those images. Real-time person tracking is enabled by combining the following seven functions (**Fig. 1**):

(1) People detection: Only people are extracted from an image (Fig. 1(a)).
(2) Attribute recognition: The gender and age group of a person are estimated.
(3) Detailed attribute recognition: Attributes associated with specific body parts are recognized. For example, a person is searched for on the basis of detailed attributes concerning their appearance, colors of clothing, and presence of personal items such as a bag (e.g., having long hair and wearing a white shirt, jeans, and sunglasses).
(4) Person re-identification: Whole-body collation is applied to determine whether the detected person is the target person (Fig. 1(b)).
(5) Trajectory recognition: The trajectory of a person walking is estimated from video images (Fig. 1(c)).
(6) Multiple camera compatibility: Functions (1) to (5) are supported even if the target person crosses the views of multiple cameras.
(7) Real-time analysis: The results of (1) to (6) can be analyzed in real time.

2.1 Achievement of whole-body collation ahead of our competitors

We achieved function (4), person re-identification, by employing whole-body collation ahead of our competitors, which makes it possible to extract people even if the person is facing backwards. Automatically

extracting features from a large number of pairs of images of people by using deep learning technology made it possible to match the images with higher precision than rule-based judgment using human-set characteristics (body type, clothing color, hairstyle, etc.).

The person-search service provided by NTT Communications called Takumi Eyes—which incorporates part of our real-time person-tracking technology—was awarded the 20th Automatic Recognition System Grand Prize by the Japan Automated Identification System Association in 2018 [3]. Winning the award indicates that this technology is highly evaluated by the market.

2.2 Results of collaboration between Panasonic Group and NTT Group

Whole-body collation is not perfect by itself. A whole-body check is difficult to perform if a person's appearance changes such as when they take off outerwear (coats etc.) that they were wearing. We devised a solution to this problem in collaboration with our partner.

A business alliance agreement with Panasonic Corporation in 2015 [4] triggered efforts to greatly improve recognition accuracy by combining our whole-body-collation technology with Panasonic's face-recognition technology [5]. Created as a result of the combined technologies was real-time person tracking, which can match people with high accuracy from camera images shot under various angles and conditions. Until now, we were not aware of any other services that combined full-body collation technology and face authentication using deep learning in this manner. At the present time, only the person re-identification is achieved by whole-body collation plus face authentication; however, the technology can be combined with additional detection functions, such as detecting the gait of a person, in response to the needs of our customers.

3. Video monitoring market forecasted to be 160 billion yen by 2030

This section focuses on the business potential of the above-described real-time person tracking using deep learning technology.

The image-analysis business is a promising sector with the highest growth rate in the AI market. It is expected to grow more than one-hundred-fold, namely, from 1.3 billion yen in 2015 to 160 billion yen in fiscal year 2030 [6]. It is expected that the market for analyzing images shot by surveillance cameras will increase.

A use case of an actual implementation is described in the following subsection.

3.1 Utilization of surveillance cameras at convenience stores: person re-identification

An example that is easy to imagine as a use case involving surveillance cameras is the use of surveillance-camera images taken in convenience stores. Having surveillance cameras in present-day convenience stores enables people to confirm what actually happened from past images after an incident or accident has occurred. A convenient function in such a case is person re-identification.

In the case of a crime occurring at a convenience store equipped with surveillance cameras, it is possible to quickly find when the perpetrator entered the store by specifying the image of the person from the video at the time of the crime and then searching from other past video images. In addition, it is possible to quickly find out if the crime was planned or impulsive by retrieving the perpetrator's past images in order to determine the history of their store visits—if they had been to the store to check it out in the past. Moreover, person re-identification is even more effective in the case of large facilities such as apartment blocks or shopping malls fitted with multiple surveillance cameras.

Scenes in television crime dramas in which a police detective spends a long time checking surveillance camera videos will surely be a thing of the past once this technology is put into service. Furthermore, utilizing this technology will eliminate oversights due to human error.

3.2 Utilization of surveillance cameras in commercial facilities: finding lost children by combining attribute recognition, color search, and person re-identification

Finding lost children is an expected use case for larger-scale commercial facilities. When shopping at department stores or shopping malls, we sometimes hear announcements about lost children. Announcements such as, "A mother is looking for her five-year-old daughter (name), who is wearing a pink dress." are commonly heard on busy weekends at commercial facilities. Such announcements, however, may no longer be necessary once this technology is put into service. The child's age is specified by a technique called attribute recognition, and the color of the child's clothes is specified and searched for via a

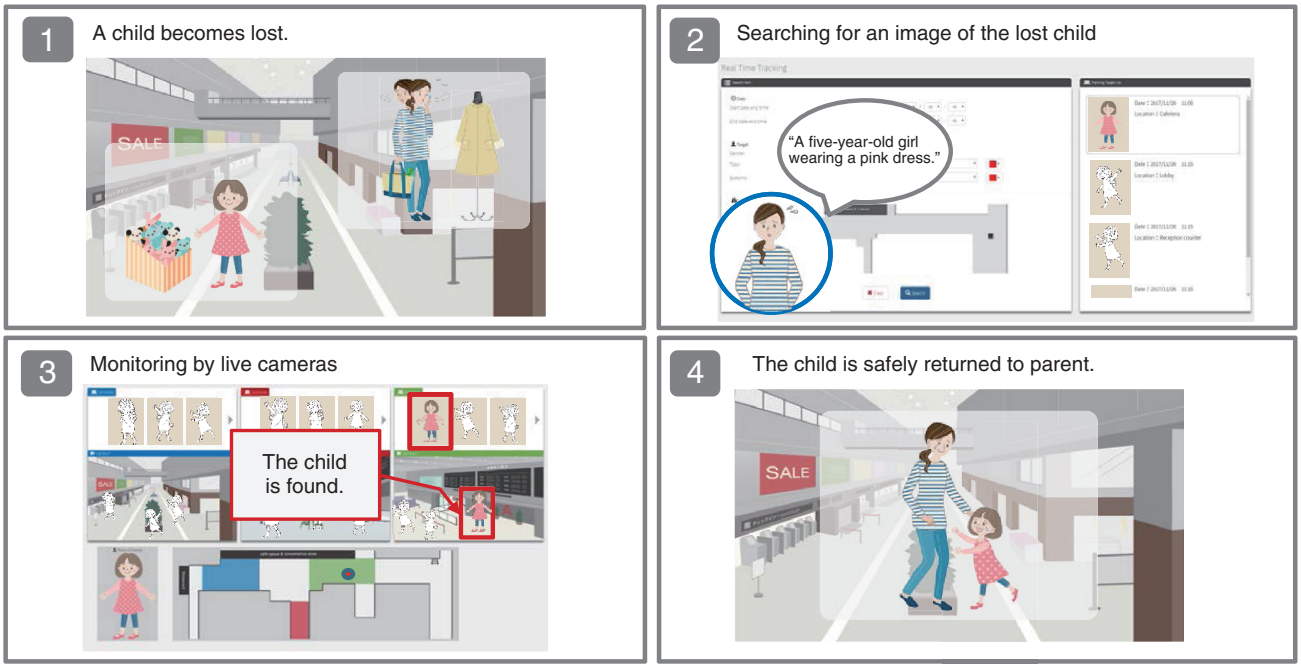


Fig. 2. Use of surveillance cameras in commercial facilities to look for a lost child.

color search. From the images captured by nearly 100 surveillance cameras, the most-recent captured image of the child can be displayed, and the lost child can be pinpointed from that image (**Fig. 2**). Such a system is likely to be essential for future large-scale theme parks and other such facilities that do not implement measures for announcing lost children.

3.3 Search for wandering citizens for municipalities

The example mentioned in the previous section is a use case concerning a lost child in a commercial facility, but it can also be applied to finding elderly people who wander away from home. With one photo of the elderly person who has wandered away, it is possible to promptly find that person from the images captured by multiple surveillance cameras operated by municipalities. The person re-identification function can find people in a much shorter time compared to manually checking surveillance-camera images with the human eye. Using this technology as a human assistant makes it possible to promptly find the wandering elderly person and thereby reduce the probability of that person being involved in an accident. The news footage of people in local municipalities, including hundreds of local police and fire fighters, simultaneously searching for elderly people who have wandered away may not be very common after this technology is put into service.

3.4 Use in marketing: combining person re-identification, attribute recognition, and trajectory recognition

This real-time person-tracking technology can also be used for purposes other than crime prevention, namely, marketing. It may also be useful for analyzing and outputting the number of visitors and their attributes in a time slot from surveillance camera images as well as analyzing the flow of customers. It will be possible to analyze data and determine how to arrange shelves and goods in a limited store space. Other applications are also possible. For example, although a logic program is required for detecting objects or people, by preparing training data and learning from it, it will be possible to extract items that customers had taken from a shelf but returned to the shelf (that is, products that a customer seemed interested in but did not purchase). With the introduction of such technology, it will be possible to obtain information that cannot be extracted from cash-register POS (point of sales) data.

4. Core technology that enables real-time processing

One particular core technology is important for establishing real-time processing. That technology is a technique for optimizing the deep-learning inference environment. This optimization technology makes it possible, for example, to detect and classify objects in video images at high speed and process them in real time. By combining the world's latest technologies listed below according to the type of deep learning processing (i.e., detect and classify) that is executed, we have increased the processing speed by more than 10 times.

- Image-analysis algorithm for detecting and matching people with high accuracy
- Reduction of parameter size
- Implementation technology for optimizing inference processing of deep learning

When the person-tracking service was introduced for the first time in the world as Takumi Eyes, the main service was searching for past images by using images captured by surveillance cameras. Since then, real-time processing has become possible by researching and combining optimization techniques for deep-learning inference environments over time.

5. Future development

After developing the video-analysis technology that enables real-time processing, we plan to focus our research on distributed processing of video-analysis technology. Specifically, we are planning to conduct research on distributed processing that enables systems (including central offices and datacenters of telecommunication companies as well as cloud services) to be constructed and the functions required at each location (edge) to be provided. We call this processing a *two-layer edge model*. We will work to make the operation of edge devices more efficient, lower in cost, and with the maximum savings of power and memory.

From now onwards, the NTT Group will continue to work towards applying its AI technology called corevo® and implementing it in society in cooperation with various partners in order to improve the lives and businesses of many customers.

References

[1] Press release issued by Gartner, "Gartner Identifies Five Emerging Technology Trends That Will Blur the Lines Between Human and Machine," Aug. 20, 2018.
https://www.gartner.com/en/newsroom/press-releases/2018-08-20-gartner-identifies-five-emerging-technology-trends-that-will-blur-the-lines-between-human-and-machine

[2] T. Eda, S. Muramatsu, K. Mikami, S. Xu, "A Practical Person Monitoring System for City Security," 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS 2018), Auckland, New Zealand, Nov. 2018.

[3] Press release issued by NTT Communications on Sept. 12, 2018 (in Japanese).
https://www.ntt.com/about-us/press-releases/news/article/2018/0912.html

[4] Press release issued by NTT and Panasonic, "NTT and Panasonic Tie Up for Innovative Visual Communication," June 17, 2015.
http://www.ntt.co.jp/news2015/1506e/150617a.html

[5] Press release issued by NTT and Panasonic on Oct. 3, 2018 (in Japanese).
http://www.ntt.co.jp/news2018/1810/181003a.html

[6] Press release issued by Fuji Chimera Research Institute on Nov. 28, 2016 (in Japanese).
http://www.group.fuji-keizai.co.jp/press/pdf/161128_16095.pdf

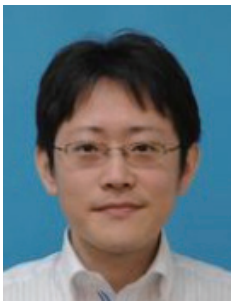
Kunihiro Moriga
Senior Research Engineer, Supervisor, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.
He received a B.S. and M.S. from Yokohama National University in 1991 and 1993. He joined NTT in 1993. His current research interests include deep learning and AI base technology.

Takeharu Eda
Senior Research Engineer, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.
He received a B.S. in mathematics from Kyoto University in 2001 and an M.S. in engineering from Nara Institute of Science and Technology in 2003. He joined NTT in 2003. His research interests include a wide range of topics related to SysML (Systems Modeling Language) such as distributed training, efficient inference runtime, scalable surveillance applications, and theories for deep learning. He is a member of the Information Processing Society of Japan and the Association for Computing Machinery.

Masashi Toyama
Senior Research Engineer, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.
He received a B.S. and M.S. in information and computer science from Keio University, Kanagawa, in 2003 and 2005. He joined NTT in 2005. His current research interests include data science and software engineering.

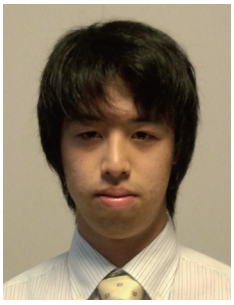
Keita Mikami
Senior Research Engineer, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.
He received a B.S. and M.S. in information and computer science from Waseda University, Tokyo, in 2005 and 2007. He joined NTT in 2007. His current research interests include data science and software engineering. He is a member of the Information Processing Society of Japan.

Yutaka Hirokawa
Engineer, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.
He received a B.E. and M.E. in computer science from Tohoku University, Miyagi, in 2003 and 2005. He joined NTT in 2005. His research interests include anomaly network traffic detection.



Yuji Yamada
Engineer, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.
He received a B.E. and M.E. in computer science from The University of Electro-Communications, Tokyo, in 2009 and 2011. He joined NTT in 2011. His research interests include data science and software engineering.

Sanae Muramatsu
Engineer, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.
She received a B.E. and M.E. in computer science from Nagoya University, Aichi, in 2011 and 2013. She joined NTT in 2013. Her research interests include deep learning and software engineering.



Taku Sasaki
Engineer, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.
He received a B.S. and M.S. from Tokyo Institute of Technology in 2014 and 2016. He joined NTT Software Innovation Center in 2016. His current research interests include attention-based deep learning and computer vision.

Shin'ya Yamaguchi
Researcher, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.
He received a B.E. and M.E. from Yokohama National University, Kanagawa, in 2015 and 2017. He joined NTT Software Innovation Center in 2017. His research interests include deep (machine) learning, particularly transfer learning, representation learning, and deep generative models.

Katsuo Inaya
Senior Research Engineer, Supervisor, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.
He joined NTT in 1995. He is an experienced engineer with a long history of working in the information technology and services industry. He has experience in the areas of enterprise software, business development, strategy, strategic partnerships, and mobile devices. His current research interests include deep learning.