

Forum Paper

Emerging AI Technologies for Smart Infrastructure

Wen-Huang Cheng¹, Jenq-Neng Hwang², Ivan Bajic³, Rei Kawakami⁴, Shiqi Wang⁵ and Jiaying Liu^{6*}

¹*Department of Computer Science and Information Engineering, National Taiwan University, Taiwan*

²*ECE Department, University of Washington, USA*

³*SFU Multimedia Laboratory, Simon Fraser University, Canada*

⁴*Tokyo Institute of Technology, Japan*

⁵*Department of Computer Science, City University of Hong Kong, Hong Kong*

⁶*Wangxuan Institute of Computer Technology, Peking University, China*

ABSTRACT

This article summarizes the panel discussion on “Emerging AI Technologies for Smart Infrastructure” organized by the U.S. Local Chapter of APSIPA on April 24, 2022. The panel discussion focuses on two topics, including the current application and future evolution of the smart infrastructure. First, several applications are discussed to analyze the key factors to build a successful smart infrastructure, which drives researchers to concentrate more on the application scenario and user demand of the smart infrastructure. Then, the future evolution direction of the smart infrastructure and potential challenges are predicted, which provides significant guidance for both the development and deployment of the smart infrastructure.

*Corresponding author: Jiaying Liu, liujiaying@pku.edu.cn. This work was supported by the National Science and Technology Council of Taiwan under Grants NSTC-109-2223-E-002-005-MY3, NSTC-112-2628-E-002-033-MY4 and NSTC-112-2634-F-002-002-MBK.

1 Introduction

[Jiaying] Rapid urbanization poses new challenges for the existing physical and digital infrastructure. The development of new infrastructure is needed in response to rapid population growth and intensive economic activities. The emerging research on smart cities offers part of the solution to these challenges and facilitates the collection of environmental data via IoT (Internet of things). Millions of sensors and devices are deployed to collect data and feedback is available after analysis. However, transmitting an enormous amount of data, preserving a complex environment, and making smart decisions in a timely manner are very demanding tasks. The recent advances in AI technology can contribute significantly to accomplishing these tasks, providing effective solutions to the smart infrastructure of modern metropolises as a system that can monitor, communicate, analyze, and act based on the data collected by the sensors. However, challenges still exist in many aspects. Intelligent sensors are dynamic and open, a concept that is still evolving. Through a new generation of communication techniques, development is proceeding in leaps and bounds. More efforts are needed to make smart infrastructure faster and more environmentally friendly. We anticipate that recent breakthroughs in various analytic solutions powered by AI will inject vitality into innovations and applications in smart infrastructure. Based on the above considerations, we call for papers for the themed APSIPA Transaction on Signal and Information Processing series. We have also tried to organize this panel around more discussion and to allow the audience to think more about these issues. We've prepared several questions for the panelists to discuss. More questions and further discussions will be made available online.¹

2 Key Factors for Successful Smart Infrastructure

The first question posed to the panel is: "What are the key factors of building a successful smart infrastructure in terms of the technologies or the applications?"

2.1 Reliability and Trust

[Ivan] Let's go back to the basics of what makes a successful product in the first place. There are two kinds of products in my mind. One kind addresses real and specific areas like water, food, health, and so on. You do not need to market them much as there is always a need out there. The other is a different kind of product for which there is no real need but which turns out to be super successful; for example, applications for smartphones and so on. We used to

¹Panel video link: https://drive.google.com/drive/folders/1kjDLznt06NrorfjDmW_S-jKKq_jlTir-h?usp=sharing.

live without them but now we feel like we can't. There is not actually a real need for these products, but there is certainly a huge market for them. When it comes to smart infrastructure, I think both principles apply simultaneously. In some cases, it might be that the health of a city depends on some of these smart applications as cities grow more crowded and polluted. Some aspects of this infrastructure meet real needs. For others, there may be a market even if we could live without them, but they may improve the quality of life in a particular way that makes people want to pay for them. So that's my view on this kind of semi-business situation. What is also needed is reliability. The more we offload decision making to technology, the more we need to trust it. And I think this is a big challenge for AI to overcome.

2.2 Information Sharing and Connection

[Jenq-Neng] My personal take on infrastructure relates to GPS navigation, as shown in Figure 1. We have been enjoying the benefits of GPS, and one notices that almost every car has it. GPS information is processed through centralized cloud services such as Google Cloud and then returned to the user. My fundamental thought on these infrastructures is that as long as everybody is willing to share the information and is also an IoT sensor carrier, then we can easily enjoy a good infrastructure. This is because while people may not know that they are data contributors, they are making the system better. This is a very simple type of infrastructure which allows everyone to benefit. Now everybody is contributing to the supply of data. If you look at the reason, what IoT sensors are now in play? Except for GPS, we are talking about cameras, radars, and lidars. More and more IoT sensors are being deployed, whether fixed surveillance camera or moving cameras on every car. With current AI technology, we can easily perform advanced techniques such as object detection, 3D localization, multiple camera tracking, and so on. With all these cameras, one can detect, track, 3D localize, and determine the speed of vehicles. Additionally, you can even track a vehicle across a multi-camera system. If you have the information, you can likely also perform object segmentation and 3D sizing using a camera (sometimes with the help of lidar or radar). In addition, AI today can even help us understand human poses in the 3D space. If you look at this kind of infrastructure, we gradually realize that the approach is simple. We should be able to allow all these infrastructures and IoT sensors to perform different kinds of tasks separately.

Take GPS navigation as a simple example. We should think about the next generation of smart infrastructure as a coordinated innovation, shared for small cities and communities, looking at all vehicles with all these infrastructure types of cameras, sensors, cars, cameras, radar, and lidar. All participants are generating data. If they are willing to share their information through mobile edge computing, which is getting more and more popular, we are almost

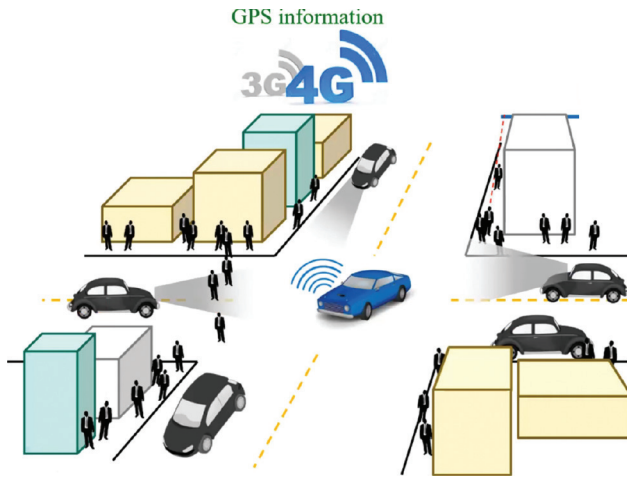


Figure 1: Smart transportation with online GPS navigation. Image source: Jenq-Neng's PPT.

ready to build these systems since we are not simply sharing simple GPS information we are sharing the whole environment. For example, predicting when another car is about to steer to the right, even before the driver can know what's going on here and may not even have a sensor. They are able to experience almost everything surrounding them as long as their vehicle has a 5G receiver.

I would ask if you propose using another concept, is there any testbed like this? I was helping this group set up a testbed in Xuchang, Henan, China. Most people say, I just need to have a very fancy center like a Tesla car or Google Waymo: I need the radar, the camera, all different kinds of sensors. But they are collecting and processing data anyway. In this testbed, even if you don't have any sensors at all, you can still enjoy all different kinds of autonomous driving aids using the street infrastructure. Cameras and lidars are found every 50 meters and sense everything happening on the road. With this capability, you just use 5G to send information to all vehicles, and everybody can experience the virtual environment. In order to be successful, you need to get all the coordinated data and perform information mining, instead of everybody just minding their own business and not communicating with each other, which is not going to provide a very useful infrastructure.

This kind of smart city concept can be also easily applied to the smart community more generally. I have also been helping the Taiwanese government to create long-term care centers because, in many Asian countries, we have more and more aging people. During the daytime, seniors may need to be sent to the long-term care center. For instance, 22 cameras are installed in

the center, and all the seniors are closely monitored. Everybody returns to the family home at night. Complete information about behaviour during this time is not available for analysis. There is no privacy issue: the issue is only specific to that person's analytical information. The whole intention is that AI should not be just simply used for nano applications.

We are also helping National Oceanic and Atmospheric Administration (NOAA), the administration tasked with monitoring Earth's atmosphere and oceans. We have been working with them for about 12 years. They want to know how many fish are in the sea and how many have been captured by the world's fishermen because they need to understand the ecological system and possibly also how it relates to global warming [1]. Given these goals, they need to equip fishing vessels to do the analysis. How many fish might be there, and of what size and what species? This takes a lot of human labor, since it ends up with the fishermen doing the job because they go out to catch fish every day. So now they are setting up a smart system, placing many cameras on the ocean surface, for example in the Gulf of Mexico or around Alaska. In addition, every fishing platform is given this kind of monitoring, relieving fishermen from detailed harvest reporting. They just need to report whatever fish they catch one by one to the monitoring infrastructure. Cameras are even placed next to the fishing platform. Based on all this information, it is possible to make complete and accurate counts, including all the sizing and species classification. Imagine we have 1,000 fishing platforms on the Pacific Ocean. We really can answer the question: how many fish are there still in the sea? This was asked, for instance, in an interview in the Wall Street Journal two years ago. So basically, the idea is that all the AI-enabled fishing platforms provided automatic reporting. How many fish are possible in that area, of what kinds and sizes? This is similar to our GPS information example: we are equipping all the fishing platforms with the infrastructure to allow us to understand what one may call the smarter ocean.

2.3 Smarter Decision with AI

[Wen-Huang] I think it's good to follow Professor Hwang because our vision is based on very similar technology and goals, but from a different perspective. Instead of so-called big data for "small" oceans, maybe I can focus on a different perspective like AI or big data for small logistics. Because we know that in the real world, we don't just have people; we have goods, and also many different business activities. Logistics is a very important area. What is logistics, precisely? It is the process of planning and executing the efficient transportation of goods from point of origin to point of consumption, including storage. What are the major functions of logistics? There are two main functions: transportation and warehousing. Transportation, specifically, means transportation management, focused on planning optimization in the execution

of vehicle usage to move goods between warehouses. Warehousing involves the management of the flow and storage of goods, as well as tracking inventory.

Why have I chosen this topic? Because the demand for quick and convenient delivery for almost everything is ever increasing. AI (e.g., driverless technology) is expected to play an important role in the future of logistics. Why do we need to employ AI for making better or smarter logistics? Consider a very simple case. In Figure 2, you can see that we have users A and B. Usually when we want to get something from user A to user B, we expect that it will be transported from user A to user B directly. But in today’s logistics reality, all the goods will be transported to a central warehouse and then distributed to the various destinations. This scenario works well in a downtown environment because distances between the users and the warehouse are very short. But you can imagine that in a larger urban area, while the different users could be very close, the distance to the warehouse could be much farther. So if we still need to transport everything from one user to the essential warehouse and then back to another user, it will be very inefficient. Just as in Professor Hwang’s presentation, we can use AI technology like driverless or tracking technologies to monitor all the goods and all the users, improving transportation planning. You can see that the logistics journey is very long from the manufacturer until the end-users; while it is a very simple and straightforward journey, at each point in the journey the so-called decision is very difficult. So we intend to rely on AI to improve performance.

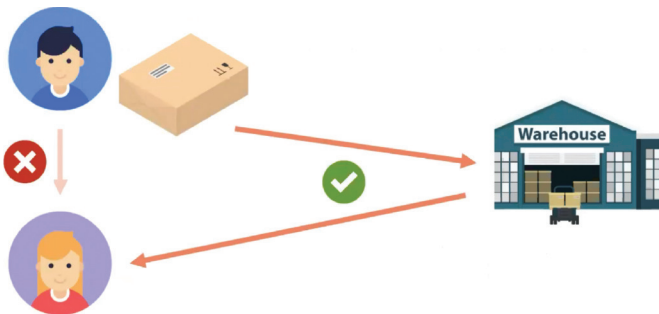


Figure 2: Transportation prediction problem using AI. Image source: Wen-Huang’s PPT.

Let me take packing as a second example. Packing is easy to understand in theory. We just place the object in a cardboard box and transport the box somewhere. But you can see that packing itself is a very challenging algorithmic problem. For example, we have paper boxes in different sizes, and we also have a set of goods. How to use the least number of boxes to pack all the goods is a challenging algorithmic problem. So you can see that at key conferences like AAAI and ICML from 2000 to 2021, there has been and continues to be a lot of research efforts [5, 6] on investigating such an apparently simple question.

In the last year, Huawei organized a competition with a focus on logistics. In this competition, they focus on three different problems. The first one is the vehicle routing problem; the approach is to try to leverage solutions to optimize the routing of different vehicles in the transportation process. The second and third problems, the splitting delivery problem and the 3D bin packing problem, are related to each other. We have a lot of opportunities ahead to build a smarter infrastructure for many issues such as Professor Hwang just mentioned for the autonomous driving and smart ocean scenarios. I would like to share other scenarios like logistics, which is also a big area.

2.4 Robustness to Real-World Problems

[Rei] I am also thinking about a lot of real-world problems. We can consider many applications of large pre-trained models to solve real problems. The great success of neural networks is based on web-crawled images. But the images that we want to deal with are specific. It depends on what kind of problems you want to solve. I have been working on ecological assessment, smart agriculture, and autonomous driving. Each of these problems has very different image statistics compared to web-crawled images. We need to solve the real-world problem, so we basically need to improve the inner robustness of the recognition component. Besides, you need to handle some anomalies that can happen in the real world, which relates to improving the reliability of the system.

Let me show some examples. Figure 3 shows one of the sample images that I have been working with. It shows a windmill. Bird species at risk of extinction can strike the turbines, leading to a potentially severe impact on the ecological system. So the aim was to count the birds in the area and ensure that windmills are not impacting the ecological system [3, 4]. As you can see,



Figure 3: Bird detection result [4]. Image source: Rei's PPT.

you have a 4K image but the birds in the camera field are actually around 20×20 pixels. And you also have MPEG artifacts. So your targets are kind of noisy moving objects. But if you use multiple images, then you can use the motion pattern to recognize those tiny things. The blue bounding boxes in the figure are the candidate regions for the birds, for the moving objects. And then the squares in red are the ones that we found. This is an image that is typical of hard problems in reality. And also, you need to deal with how many cameras you want to install in these energy systems. And you also need to think about saving energy as much as possible so that you do not need to have a massive GPU around the top lines, although the energy itself is not a problem in this environment.

I was also working with a small company that wanted to try to find embryos within egg shells. They wanted to count those embryos that had developed eyes, so that they can estimate the number of embryos with eyes they can harvest from the sea. But when you see the real problems that you need to solve, then you feel that they are very specific and very different from the available techniques. It's not clear what is the optimal way to solve these problems. But actually, the large models are really easy to transfer to these individual problems. Still, we need to think about the optimal way to solve these problems so that many people can reap some kind of reward from the large models.

Figure 4 shows another scene. It is an extended area of land with many cows. We mark the cows so that you can count how many there are in this large area and make sure that no cows have suffered illness or accidents [2]. I was very surprised at how robust those pre-trained models are when I tested



Figure 4: Successful cow detection result [2]. Image source: Rei's PPT.



Figure 5: Failed cow detection result [2]. Image source: Rei’s PPT.

them with these real-world images. But I also found that if the angle is slightly different, then suddenly the recognition can be unreliable, because bushes and cars are counted as cows, as shown in Figure 5. So there are still a lot of techniques to develop to make your system robust to the environment for solving real-world problems. And it’s very necessary to avoid any anomalies.

2.5 Organization and Explainability

[Shiqi] To build smart infrastructures, the first thing to say is that it’s quite different from traditional video coding in relation to the human recognition system. From my background and perspective, the whole pipeline should be driven by organization toward the automated utility, which means that the final product is quite different. We need to create this kind of organization. For example, when you are building a smart help application, low delay is very important. But if you are promoting the application scenario, I think all pipelines should be optimized toward this automated utility. The second thing that I want to mention is the capabilities of the organization, which are also very important, because we have different acquisition conditions in real-world scenarios. And also, we have different bandwidths and even levels of decoding complexity for these images and videos. There are different kinds of domains. If we train a model in one specific domain and apply the model to another kind of domain, this will create a lot of gaps in the training and testing in these smart applications. We have some techniques like domain adaptation that could be very useful in this kind of scenario.

The last thing that I want to mention is the model. We see that “smart” infrastructure must be genuinely smart, relying on the particular kind of models of the learning system. So basically, the models should be explainable. I think

most of the models, as you may know, are black boxes. But how to make the models robust, how to read models, and how to be convincing with respect to different applications? I think that would be very important. So that's my understanding of this problem regarding key factors for creating smart infrastructure. We need to organize towards ultimate utility, and we need to have strong organizational capability. And also, the models or the entire system should not be a black box but explainable, robust, and trustful. From my point of view, they will be very important in this whole area of smart infrastructure.

3 Successful Examples of Smart Infrastructure

The second question posed to the panel is “Are there any examples of successful smart infrastructure or interesting applications realized via a smart infrastructure?”

3.1 *Smart Manufacturing*

[Ivan] I think we've heard a lot of good examples already. I think those are all the sort of emerging applications where there are still a lot of issues to be resolved, and much research to be done. In some sense, the most established applications may have been in smart manufacturing. Part of that is it's a well-controlled environment. You build a production facility from the ground up; you can minimize unexpected events. The general public is not allowed in, so the variability of conditions under which this smart infrastructure needs to work is reduced. And maybe that's part of the reason why I would say those kinds of applications, in my mind at least, seem pretty well advanced and established. Anything that has to deal with some sort of unpredictability, such as smart homes, smart neighborhoods in cities, and smart hospitals, still has many issues. Things are emerging, but there is still research to be done to fully realize the potential correct technology. And we have heard some excellent examples already in other applications.

3.2 *Widely Connected Edge Devices*

[Jenq-Neng] I can elaborate a bit more on the potential examples. If you think carefully about why we are talking about smart cities, smart transportation, smart communities, and smart manufacturing nowadays, it is because we have been more and more connected to IoT sensors. We also have a very good communication infrastructure, 5G and 6G, and very powerful computing power, GPU and TPU. Finally, we also have the capability to do this with really effective control and feedback. With everything combined, we can call it smart. Otherwise, if we proceed individually, we have sensors all the

time; we have communication from 2G, 3G, and 4G all the time; we have computers all the time. But now for the first time we can close this route from sensing, communication, computing, and feedback. Because of that, I would say there will be tons of possible applications popping out. The reason is lost data, corrected IoT data; no matter what you need, there's nowhere are you cannot absorb so much data, and it's just a lot. You just use all the data without any kind of information aggregation. Whenever we do information aggregation, no matter whether you use traditional machine learning or deep learning, it's getting smarter and smarter. And that's what we are talking about. Everything is smart. For example, Rei was talking about smarter agriculture. Many of my friends are also working on that. The smart health system, smart transportation, smart ocean, and smart city stuff. In the next five or ten years, smart infrastructure will be everywhere. I am looking forward to seeing all of that.

[Wen-Huang] I agree with the comments from Professor Hwang. I think we should not just have stand-alone sensors. We need to have all the sensors coupled together to make better decisions. Especially, in some sense, the human element is the biggest obstacle in the environment. Human behavior is sometimes very difficult to predict, so maybe we can try to use IoT devices with different technologies to automate tasks and collect data, and then leverage the data to make better decisions. Just like the comments from Professor Hwang, in the next five to ten years, there will be a lot of successful smart infrastructures since all the IoT devices can be interconnected.

3.3 Online Shopping and Ordering

[Rei] I agree with all three professors. But if I can think of any example of good smart infrastructure, it is probably Amazon which I use daily. It's pretty amazing in terms of how fast the products are delivered. Also, Uber. I frequently order food to feed my kids. Also, Google. I cannot travel without Google. All those big tech companies are really smart. They can aggregate information and solve privacy issues. But if more and more sensors are surrounding us, then we need to think about the privacy issues. Take Amazon speakers: if they're grabbing information by listening to your family, it might raise privacy concerns since it would be no surprise that researchers and engineers can listen to your conversations. So, we need to deal with privacy issues in the future. And also, health care and child care are areas that need human labor. Whatever can be installed without privacy issues will be very helpful for everyone. I hope those techniques will be available soon.

4 Technological Evolution of Smart Infrastructure

The third question posed to the panel is “What technological evolutions do we expect to see in smart infrastructures in the next five to ten years?”

4.1 *Relevant Information Extraction*

[Ivan] This has already been touched on by a few speakers. To summarize, I would say it will involve more sensors, higher connectivity, higher bid rates, and a lot more. I think there are many applications yet to be imagined, so I would let the students be more creative. But definitely, I think understanding of the environment of the sensor and extracting relevant information from it will be important in making more and more autonomous decisions. That seems to be the trend of where the research is going, and deploying that would require privacy handling, trust, and explainability. I wish I knew what we will be working on in the next ten years, but at the moment, we seem to be following these trends to see where they go.

4.2 *Zero-Shot Learning and Incremental Learning*

[Jenq-Neng] These days, when we say we are doing AI and deep learning, we are very much doing big-data, small-task type of applications. We use millions of data elements to try to modify a very specific task. And everybody knows it may not work because what if the sensor fails for some reason? You can just upgrade the IoT sensor. People say all these are cameras or environmental sensors: just upgrade to the next generation. Are we going to recollect all the data and do everything from scratch? That’s not going to work most of the time. What can we do and how should we do it? Incremental learning? Or the aforementioned unsupervised attention domain? We need to have that capability all the time. The most important part is that we hope everything we are going to do is zero-shot, which means no new training data is needed and new concepts can be automatically learned. This is becoming the most popular solution: multi-modality of training algorithms, which are making sense right now. For example, if you have a model that has learned from a set of visual data, next time, when you see something new from images unseen in the training data, the model should be able to recognize it without retraining with new training data. This kind of zero-shot, incremental learning or unsupervised domain adaptation need to be much more mature before we can enjoy any kind of smart infrastructure.

4.3 *Centralized Data Center*

[Wen-Huang] I guess I might not be able to foresee the evolution of the field in the next five to ten years, but I would say that I see the revolution just getting

started. For example, today, many businesses and industries are talking about digital transformation. Digital transformation means that we try to transform the data collection approach from the traditional methods, especially those where humans manually collect or organize all the data. Setting up centralized data centers, whereby we can develop some automation and predicted analytics, has been attempted. I believe that based on these processes, we can gradually see more creative and interesting applications.

4.4 High Accessibility and Generalization

[Rei] It's a tough question to answer, but let me try. One thing that we are sure of is that IT technical technologies can be installed for visually or auditorily impaired people, giving them good navigation systems that allow them to walk around freely. Those things are probably happening very soon. I think technologies nowadays can be personalized very well. So, I'm kind of sure that those with a disability will be able to work and move around more freely than before.

Another thing is that cars might become infrastructure elements as well. You don't own a car, you just pick one up, get in and drive it to your destination, and then walk away. Perhaps cars are moving all the time and users just pick them up and for specific transportation. That kind of future might be possible.

4.5 Unlocking the Data Value

[Shiqi] I think in five to ten years, we can expect to see data value fully unlocked. Actually, from my background in video compression and representation, a lot of visual data nowadays are utilized. This is actually what we are now expecting to see. We believe each piece of data has its value. So basically, I think we need to unlock the value of the data. To use the data, we must effectively represent it. I know that Ivan is working on this area. How to represent this data to make it smaller and improve the performance is very important.

5 Challenges and Opportunities for Smart Infrastructure

The fourth question posed to the panel is "What are the challenges and opportunities for building either smart infrastructure or its applications?"

5.1 Scale and Efficiency

[Ivan] I think one of the challenges is how to scale up. We've heard that billions of sensors are going to be deployed. Each sensor needs to communicate

with the others. But when everything is communicating with everything else, we get an exploding combinatorial problem. Providing the bandwidth and the communication infrastructure needed may be one of the challenges.

Another challenge is the full mining of relevant information from the captured signal. When we think of data, there is noise and information in the signal captured by a sensor. With information itself, there is Shannon information, but there is also information relevant to a particular task. The same signal may have different amounts of information as relevant to different tasks. Effective communication is transmitting only what's needed. This requires us to better understand what is relevant information in the signal. It's not simply the signal minus the noise level. What remains could still be irrelevant.

5.2 *Privacy and Security*

[Jenq-Neng] Again, while some have already been mentioned, I would like to re-emphasize the major challenges I see, especially when we are talking about infrastructure-type applications. "Infrastructure" means that there will be a lot of aggregation of information and collection of all different kinds of data. My first concern would be privacy and security. For example, everybody is aware of privacy, which is a different technology now that can be used easily. That kind of audio or video can be easily generated. Creating a fake video of someone doing something they didn't do is easy nowadays. This is a crime but it can now be done easily. Something like that is going to be terrible, because we can easily create, for example, all these Russia-Ukraine "worlds." We see all different kinds of fake videos being distributed all the time.

If fake information can fly around without restriction, all smart infrastructure would be in trouble, not to mention our privacy since a lot of data relates to our personal information. Unfortunately, our data are still sent to the cloud server and analyzed there, making our privacy more vulnerable. That's my concern.

5.3 *Pinpoint from the Current Industry*

[Wen-Huang] I think infrastructure is a kind of service architecture in support of real-world tasks. In terms of building up smart infrastructures, I would say that we have to know the pain point from the specific industry or task. The issue could manifest as a classic chicken-and-egg dilemma if we aim to integrate a solution addressing a specific pain point into our current infrastructures, thereby re-encoding the very problem we are trying to solve. The infrastructure cannot be too advanced, because it needs to be made under current limitations. On the other hand, if we just try to develop next-generation infrastructures,

people in the particular industry might complain about the architecture being too advanced and have no idea how to use it. My concern is effectiveness in the context of real-world problems.

In summary, I would say that how to know the real pain point, for example from the industry, and considering the solution to certain pain point into the building of infrastructures could be one of the challenges.

5.4 Regulation and Rule

[Rei] I think most of the important challenges have already been discussed by the three professors preceding me. What I am thinking about, for example, is when you construct a large bridge and you need to check to a phrase period. This probably requires regulations to check off. While problems that all the others have provided are very important, building some kind of regulations for smart inspection systems is also crucial. If you are doing smart inspections, then one kind of crack is at a severity level of one and another kind is at severity level two. These things can be hard to quantify, and regulations are also difficult to build. You need to work hard with all the practitioners for inspections.

5.5 Standard and Capability

[Shiqi] Basically, I agree with all the comments and suggestions on this topic so far. One point I would like to add is that maybe all devices will be smart. We have the smart edge, smart cloud, and so much front end, but how to divide the whole task into different kinds of devices, and how to assemble them as a whole pipeline is a kind of design philosophy. Therefore, I think this may require some kind of successful standard. For example, MPEG is working on some standards to make this kind of infrastructure as efficient as possible. Also, dividing this task into different kinds of devices based on the computing power they have would be very important.

6 Questions and Answers

Questions from the audience and responses from the panelists are presented in this section.

6.1 Competition between Academia and Corporate

First question: Are there breakthroughs that come from academia to impact the corporate community rather than just being a component that fits into this big map?

[Jenq-Neng] Think about this. It's not going to be easy to compete with Facebook, Google, Amazon, Nvidia, and so on. My students all go to these companies for summer internships. One person can sometimes use 64 GPUs at the same time, while there are only about 10 GPUs in my whole lab and they are shared by 10 students. From another perspective, that's also why if a university professor wants to do data collection or architectural training, it's not going to be easy to succeed. However, I realize that you are more likely to succeed if you have a domain-specific type of application. You have those guys who don't try to say, I want to compete with lots of big companies with benchmark types of data, including ImageNet and KITTI. A specific dataset is a good task to work on. Just like what I have been working on with NOAA, we have 200 species of fish. The dataset has a long-tail type of distribution, and the domain shift occurs every year. Similarly, there are data from manufacturing, agriculture, medical fields, and so on. Specifically, with medical data, you need fairly specific domain knowledge. I think academia can still survive quite well. Also, those big tech companies still want our students to help, because they cannot do those kinds of benchmark data that keep on taking a lot of parameters and developing the backbone architecture. After all, whenever you deal with data from a specific domain, you need a specific type of loss function and updating rule. It turns out that academics still survive better. Don't worry.

[Ivan] They have the facilities and they have the data. Therefore, in these types of competitions, it may be challenging for a university lab to survive. However, one advantage in academia is that you don't have to worry about quarterlies and things like that. You can think more about the future and encourage students to be more creative without constantly worrying about the product. I think that could be one aspect here. Another aspect is explainability. Companies might not want to put effort into things that don't have concrete values and are not well defined yet, but I think academia has a lot to say about those kinds of things. And definitely, companies need employees, so I think universities are still going to be in business. In fact, I think probably all of you have heard that people are complaining that we can't hire enough people unless skills like traditional compression are taught. I'm still optimistic. Let's play that way.

Biographies

Wen-Huang Cheng is a Professor with the Department of Computer Science and Information Engineering, National Taiwan University (NTU), Taipei, Taiwan, where he is the Founding Director with the Artificial Intelligence and Multimedia (AIMM) Research Group. Before joining NTU, he was a Distinguished Professor with the Institute of Electronics, National Yang Ming Chiao Tung University (NYCU), Hsinchu, Taiwan, from 2018 to 2023 and led the Multimedia Computing Research Group at the Research Center for Information Technology Innovation (CITI), Academia Sinica, Taipei, Taiwan, from 2010 to 2018. His research interest includes artificial intelligence, multimedia, computer vision, machine learning, digital transformation, and financial technology.

Jenq-Neng Hwang is currently the International Programs Lead in the ECE Department. He is the founder and co-director of the Information Processing Lab at the University of Washington. He has won several AI City Challenges awards and received many paper and society awards. He has written more than 400 journals, conference papers, and book chapters in the areas of machine learning, multimedia signal processing, computer vision, and multimedia system integration and networking. Dr. Hwang is a fellow of IEEE.

Ivan Bajic is a Professor of Engineering Science at Simon Fraser University, Canada, and a Co-Director of the SFU Multimedia Laboratory. His research interests include signal processing and machine learning with applications to multimedia processing, compression, and collaborative intelligence. He received many paper awards, like ICIP 2019. He is currently the Chair of the IEEE MMSP Technical Committee. He was an Associate Editor of IEEE TMM and IEEE SP Magazine. He has contributed to JPEG AI and MPEG-VCM standardization efforts.

Rei Kawakami is an Associate Professor at the Tokyo Institute of Technology, Tokyo, Japan. She was also appointed as a Senior Researcher at Denso IT Laboratory in 2020, doing research related to autonomous driving for two years. Her research interests are in computer vision and image processing.

Shiqi Wang is currently an Associate Professor with the Department of Computer Science, City University of Hong Kong. He has proposed more than 50 technical proposals to ISO/MPEG, ITU-T, and AVS standards, and authored or coauthored more than 200 refereed journal articles/conference papers. He received the Best Paper Award from IEEE VCIP 2019, and ICME 2019. He serves as an Associate Editor for IEEE TCSVT. His research interests include video compression, image/video quality assessment, and image/video search and analysis.

Jiaying Liu is currently an Associate Professor and Boya Young Fellow with the Wangxuan Institute of Computer Technology, Peking University, China. Her current research interests include multimedia signal processing, compression, and computer vision. She has served as a member of Membership Services Committee in IEEE Signal Processing Society, a member of Multimedia Systems & Applications Technical Committee (MSA TC), and Visual Signal Processing and Communications Technical Committee (VSPC TC) in IEEE Circuits and Systems Society. She received the IEEE ICME-2020 Best Paper Awards and IEEE MMSP-2015 Top10% Paper Awards. She has also served as the Associate Editor of IEEE Trans. on Image Processing, IEEE Trans. on Circuits and Systems for Video Technology. She was the APSIPA Distinguished Lecturer (2016–2017).

References

- [1] T.-W. Huang, J.-N. Hwang, S. Romain, and F. Wallace, “Fish tracking and segmentation from stereo videos on the wild sea surface for electronic monitoring of rail fishing,” *IEEE Transactions on Circuits and Systems for Video Technology*, 29(10), 2018, 3146–58.
- [2] W. Shao, R. Kawakami, R. Yoshihashi, S. You, H. Kawase, and T. Naemura, “Cattle detection and counting in UAV images based on convolutional neural networks,” *International Journal of Remote Sensing*, 41(1), 2020, 31–52.
- [3] A. Takeki, T. T. Trinh, R. Yoshihashi, R. Kawakami, M. Iida, and T. Naemura, “Combining deep features for object detection at various scales: finding small birds in landscape images,” *IPSSJ transactions on computer vision and applications*, 8(1), 2016, 1–7.
- [4] R. Yoshihashi, R. Kawakami, M. Iida, and T. Naemura, “Bird detection and species classification with time-lapse images around a wind farm: Dataset construction and evaluation,” *Wind Energy*, 20(12), 2017, 1983–95.
- [5] H. Zhao, Q. She, C. Zhu, Y. Yang, and K. Xu, “Online 3D bin packing with constrained deep reinforcement learning,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.
- [6] H. Zhao, Y. Yu, and K. Xu, “Learning efficient online 3D bin packing on packing configuration trees,” in *Proceedings of the International Conference on Learning Representations*, 2021.