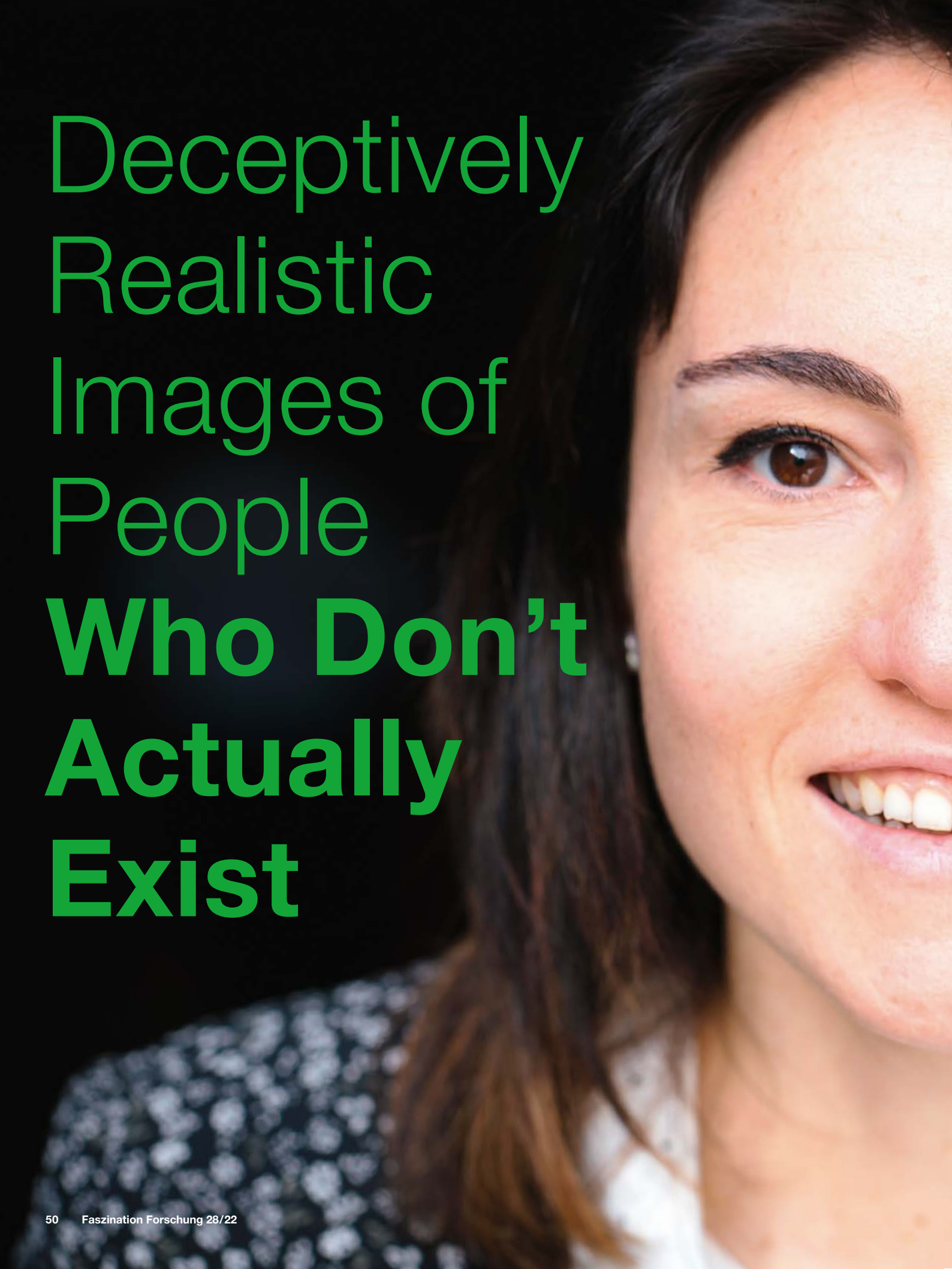
Deceptively Realistic Images of People **Who Don't Actually Exist**



Neural nets in self-driving cars require vast quantities of images of road traffic to learn how to recognize people. However, using these recordings requires the consent of the people recorded in them – which would be an enormous undertaking. A research group led by Prof. Laura Leal-Taixé has found intriguing solutions to this problem.

Link
dvl.in.tum.de/team/lealtaixe

Picture credit: Magdalena Jooss, L. Leal-Taixé

Gesamter Artikel (PDF, DE): www.tum.de/faszination-forschung-28

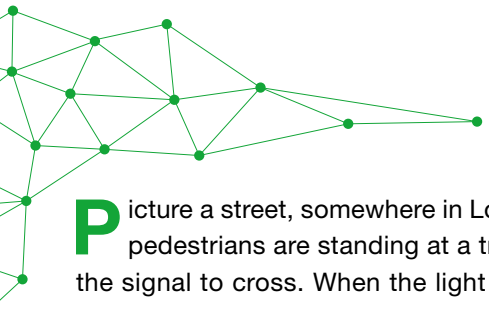
Täuschend echte Bilder von Personen, die es nicht gibt

D

Es betrifft alle Einsatzbereiche von Künstlicher Intelligenz: Neuronale Netze, seit einigen Jahren die unersetzlichen Arbeitstiere der KI, erlernen ihre Aufgaben durch Training mit ungeheuren Mengen von Daten. Oft sind es Bilder von Personen, etwa wenn ein autonomes Fahrzeug lernen soll, Menschen zu erkennen und ihnen auszuweichen. Doch solche Daten sind sensibel, die abgebildeten Menschen müssten um Erlaubnis gefragt werden. Prof. Laura Leal-Taixé erforscht mit ihrem Team Verfahren, die in der Lage sind, datenschutztechnisch unbedenkliche Bilder von Menschen zu generieren. Sie verfolgt dabei unterschiedliche Zugänge: Einerseits werden in Abbildungen von Menschen die realen Personen durch computergenerierte Personen ersetzt, die in der Realität nicht existieren. Andererseits können auch Bilder verwendet werden, die zur Gänze am Computer entstanden sind. Zum Generieren verblüffend realitätsnaher Bilder von Fußgängergruppen nutzt die Forscherin das populäre Videospiel Grand Theft Auto V. □

Prof. Laura Leal-Taixé

leads the Dynamic Vision and Learning working group at TUM. The Barcelona-born computer scientist completed her Master's thesis in Boston and gained her doctorate at Leibniz University Hannover before taking up research positions in Michigan and Zurich. She transferred to TUM in 2016, winning the Sofja Kovalevskaja Award one year later and the Google Faculty Award in 2020. Leal-Taixé exercises to unwind and has discovered bouldering since arriving in Munich.



Picture a street, somewhere in Los Angeles. Groups of pedestrians are standing at a traffic light, waiting for the signal to cross. When the light changes, they all set off, weaving and sidestepping each other as they walk to the other side. At first glance, these appear to be entirely normal people living normal lives in a big city, wearing different clothes, of different ages, different genders and different ethnicities. They don't seem to notice that they're being observed.

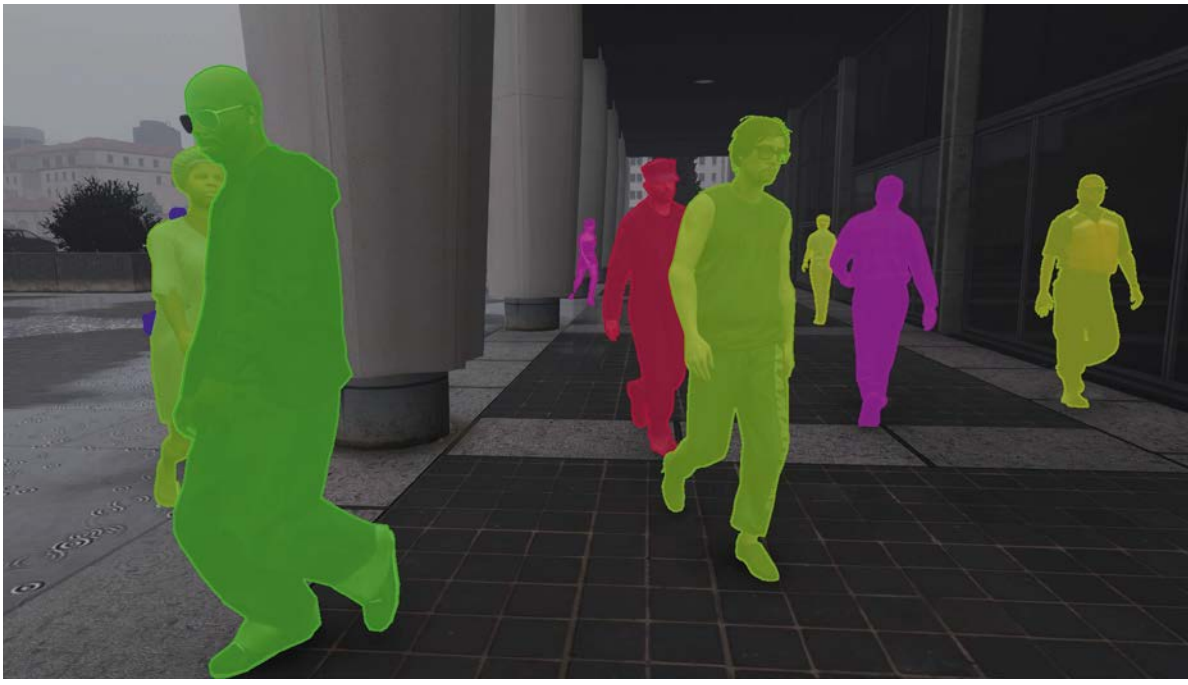
The person monitoring their activities on the screen is computer scientist Laura Leal-Taixé. Together with her team, she is putting together an extensive dataset that will enable intelligent computer programs to learn to recognize and track people in video images – a key ability for self-driving cars. This requires immense volumes of detailed images of realistic urban environments – and the images that flicker on Leal-Taixé's screen are also of high quality. Even the faces are clearly recognizable. But what about the personal rights of the people in these images? None of them have given their consent to be recorded. Given that training neural nets requires large groups of people, obtaining declarations of consent from all of them would be a mammoth logistical undertaking. And what would the researchers do if just a single person refused? Yet, Leal-Taixé has not let such quibbles affect her work as, upon closer examination, all is not as it first appears. The images on her monitor are not real; instead, they come from one of the world's most popular video games. Her team employed methods from the gaming scene to make the game useful for scientific purposes. And, in doing so, they found a surprising solution to one of the trickiest problems in computer science.

Conflict between data protection and technology

Image recognition is Laura Leal-Taixé's specialist field – and specifically the recognition of people in camera images. This technology is highly sensitive. "These methods could be used for nefarious purposes," she emphasizes. "However, sectors such as the automotive industry urgently need these capabilities." The researcher is not willing to simply leave this conflict unresolved. "It must be possible to find a balance between the two aspects. We want to solve this problem," says Leal-Taixé. ▷

Picture credit: Magdalena Jooss

△ **Images of people in video games** are suitable datasets for training neural nets. The research team has produced a massive set of training data from such images.



▷ **Video game images** already provide the information needed to track pedestrians and infer their pose.

Exactly how to reconcile the process of generating enormous quantities of data with the personal rights of data subjects is a key issue in the field of artificial intelligence. There is no shortage of voices issuing warnings, predicting that Europe will fall behind technologically if regulations are not relaxed for research projects. Many argue that research interests should be given priority over personal rights.

This is exactly where Leal-Taixé and her research group have focused their work, highlighting different ways to generate the requisite training data for algorithms without infringing on individuals' privacy.

Deceptively realistic video games

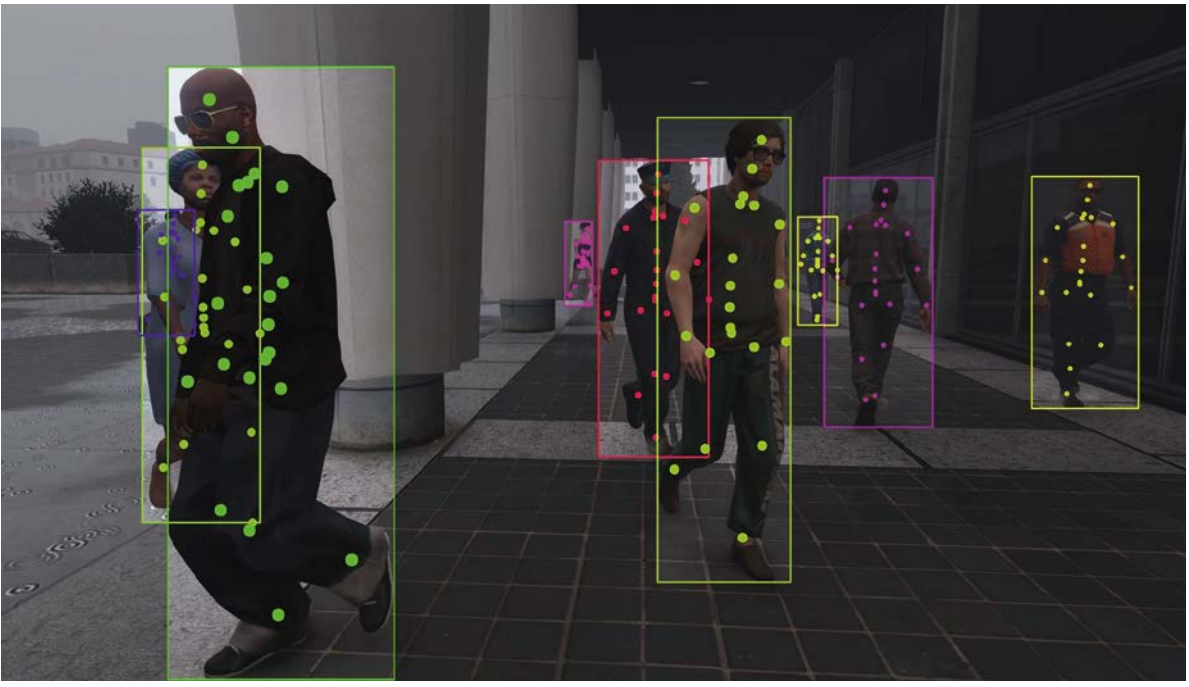
One solution is to create entirely computer-generated training images. Rather than programming these images from scratch, the researchers can draw on a pop culture phenomenon. The video game industry has been growing rapidly for many years, so much so that its sales figures have now eclipsed those of the global film industry. Realism is a crucial selling point and, as computer hardware becomes more powerful, is also becoming increasingly achievable. Past research projects have already demonstrated that images of people in video games are suitable for training neural nets. Leal-Taixé's team has now produced a massive set of training data from video game images.

Her work has focused on the popular game Grand Theft Auto V, often known simply as GTA V. Set in a major city modeled on Los Angeles, the game features impressive realism and allows players to move around freely. Beyond its realism, however, GTA V presents another vital advantage: it has become a firm favorite of "modders" – people who create modifications for video games as a hobby. Ready-made software tools enable modders to intervene in games, including by placing people at specific locations.

Useful background information

Not only does this approach avoid issues concerning personal rights, it also offers several further advantages. In real images, for example, it is important to identify the information that is actually in images, such as the real coordinates of the people in the images, in order to compare them with the results of the algorithms being trained.

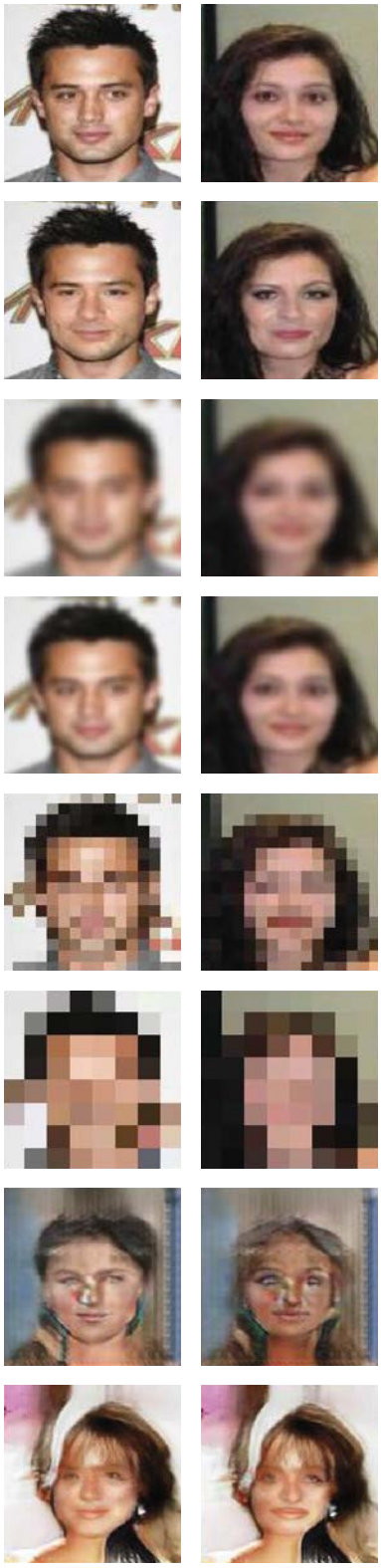
▷



Picture credit: Fabbri et al. 2021: MOTSynth: How Can Synthetic Data Help Pedestrian Detection and Tracking? arxiv.org/abs/2108.09518 (retrieved 13.6.2022).

Left: Segmentation masks

Right: Bounding boxes and keypoints of the person's skeleton describe the pose



This is known as determining the “ground truth” and can be very time-consuming. In computer games, however, the underlying information is always available and can be accessed directly.

But is synthetic image data generated by an entertainment system genuinely realistic enough to train algorithms for use in road traffic? Leal-Taixé warns against unreasonable expectations: “If you only use synthetic data in your training, the performance with real image data will not be perfect. Neural nets are very sensitive when it comes to image texture.” One way to remedy this shortcoming is to follow video-game image training with training using real-life images that have been anonymized to protect the personal rights of the people in them.

Blurring is not enough

Anonymizing real photos is another research focus of Leal-Taixé’s team. The most common way to anonymize personal data is to blur or pixelate faces. However, Leal-Taixé notes several issues with this approach. “If you only anonymize a person’s face, they might still be identifiable by other physical characteristics. We also can’t train an algorithm to identify people using images where faces are missing. The images need to be as realistic as possible,” she says.

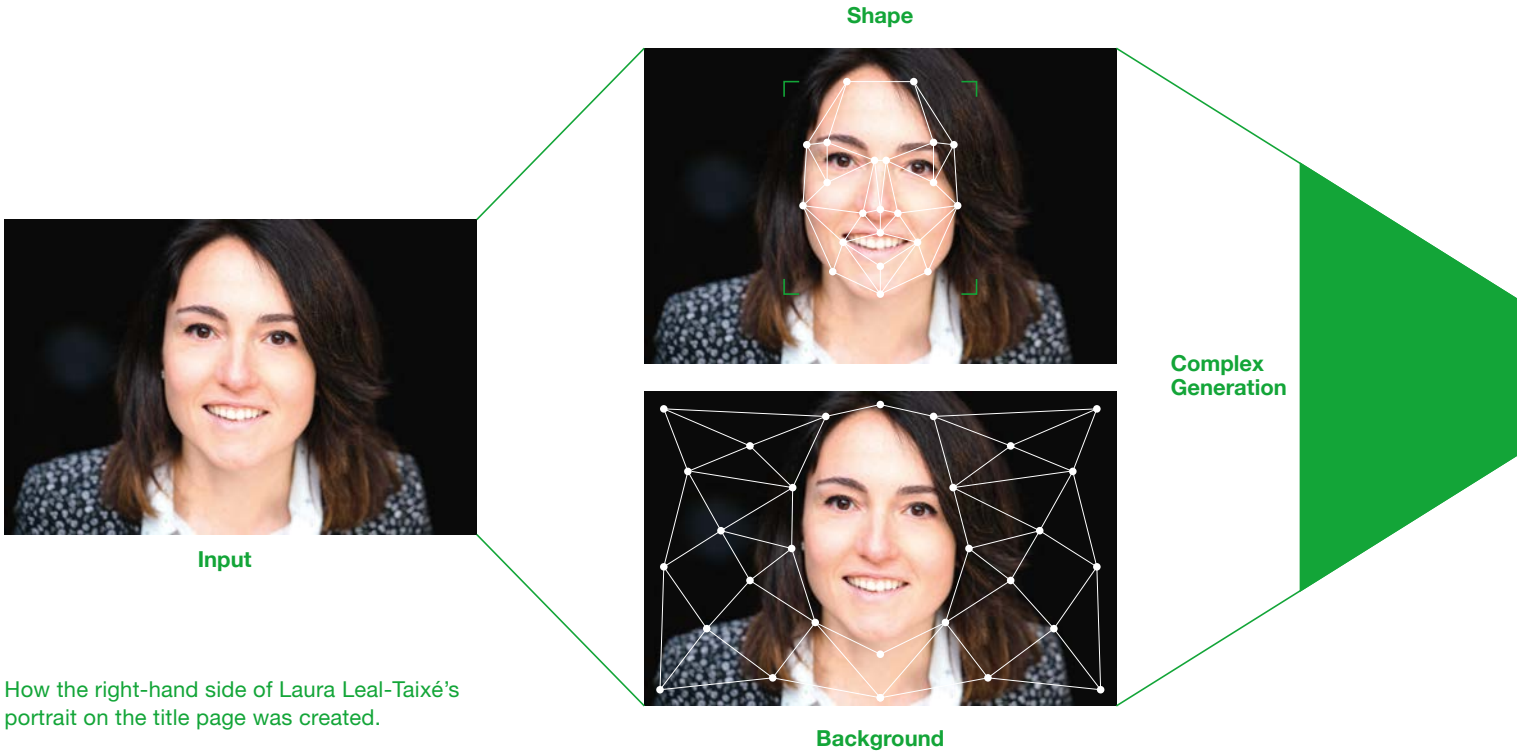
This is why, for some time, there have been attempts to distort images of people so that they can no longer be recognized by their faces. Yet even this is not always effective: it has been demonstrated that, in many cases, distorted faces can be reconstructed in full. “Deepfakes”, which have gained attention as something of an Internet phenomenon and involve replacing a person’s face with somebody else’s with remarkable precision, still do not solve the problem – as the owner of the new face also has personal rights. So, the Munich-based research group has been forced to go one step further. They have shown that it is possible to remove people from images and replace them with entirely computer-generated individuals who look real but do not exist in reality. ▷

Anonymization methods in comparison. From top: original; CIAGAN process developed by Leal-Taixe's team; blur (17 x 17 and 9 x 9); pixelation (16 x 16 and 8 x 8); image-to-image translation (Pix2Pix and CycleGAN)

“If you only anonymize a person’s face, they might still be identifiable by other physical characteristics. We also can’t train an algorithm to identify people using images where faces are missing. The images need to be as realistic as possible.”

Laura Leal-Taixé

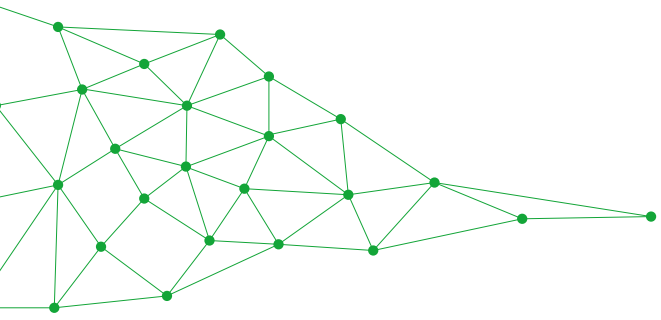
Picture credit: Maximov et al. 2020: CIAGAN: Conditional Identity Anonymization Generative Adversarial Networks. arxiv.org/abs/2005.09544, Liu et al. 2015: Deep Learning Face Attributes in the Wild. mmlab.ie.cuhk.edu.hk/projects/CelebA.html (retrieved 13.6.2022).



How the right-hand side of Laura Leal-Taixé's portrait on the title page was created.

Privacy-preserving

fake images ...



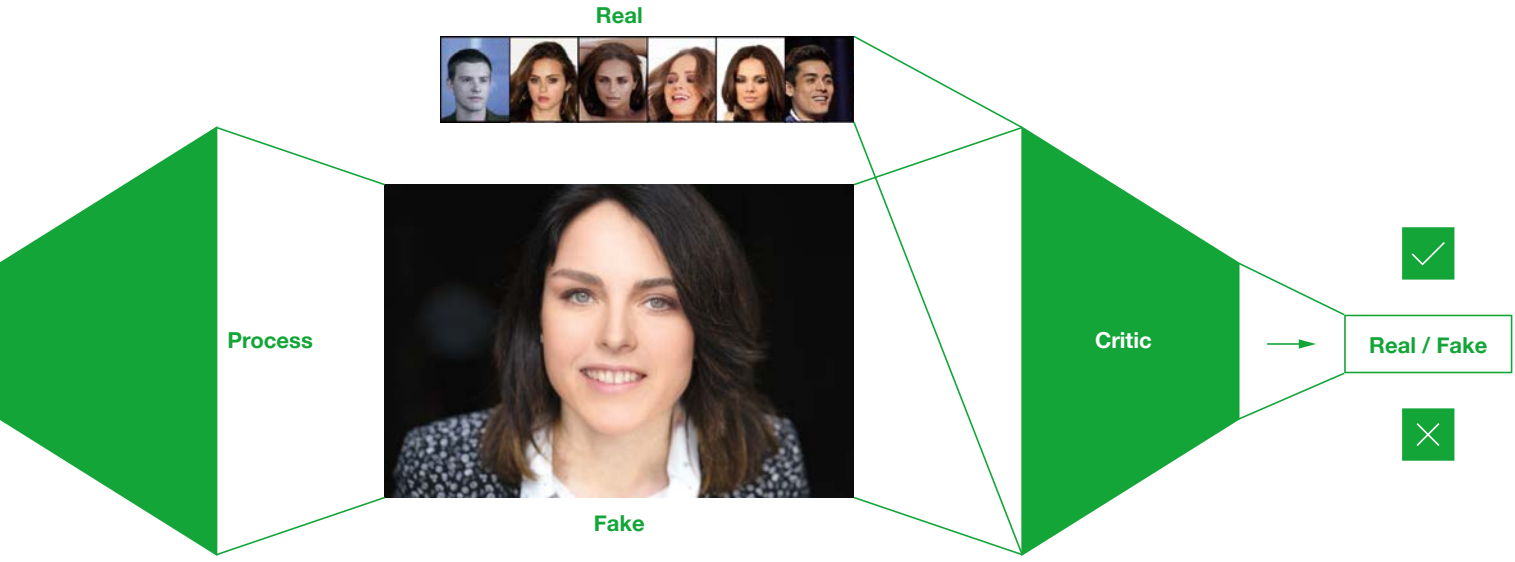
Also suitable for video footage

But is that enough to support smart mobility applications? Ultimately, moving images are subject to special requirements. If the researchers replaced a person frame by frame, it could result in jerky, unrealistic movements. "We need image stability across multiple frames in a video so that we can train algorithms to track people," says Leal-Taixé. This is possible by first focusing on determining so-called "landmarks". "In terms of faces, these include the tip of the nose, the eyes and the mouth," the researcher explains. A person's posture is another such landmark. Identifying landmarks makes it possible to recognize and imitate and person's pose, thereby preserving a realistic movement pattern in moving images.

While one neural net generates the fictitious people, a second neural net monitors the success of the process by trying to identify the people. If it cannot identify them, it serves as proof of successful anonymization.

Although this research is still at an early stage, this appears to be a highly promising approach. Leal-Taixé's group has successfully generated realistic images of people who do not really exist.

Picture credit: Magdalena Jooss, L. Leal-Taixé, Maximov et al. 2020: CIAGAN: Conditional Identity Anonymization Generative Adversarial Networks. arxiv.org/abs/2005.09544 (retrieved 13.6.2022). Graphics: ediundsepp (source: TUM)



Anonymizing faces in a way that they are not identifiable but detectable for tracking: The input image together with its landmark (eyes, nose, mouth) and shape are fed into a complex image generator to produce a fake image. This image is then scrutinized in a separate process to ensure that real persons cannot be identified.

Young working group

Leal-Taixé relies on a young team in this project. Shortly after transferring to TUM, one of her projects was recognized with the Alexander von Humboldt Foundation's Sofja Kovalevskaja Award. The € 1.65 million endowment enabled her to establish a working group. She believes that the student pool and open environment have offered the ideal conditions for her work. "The students are exceptionally well prepared and we would love to be able to supervise more theses."

As it happens, when Leal-Taixé shuts down her computer and heads home for the day, she leaves the video games in the office. Her interest is purely professional, though that wasn't always the case. She also enjoys playing open-world games like GTA in her leisure time – a secret vice, she admits. Since the birth of her daughter, however, she has little time for that. Instead, she prefers to let algorithms play the games, so they can learn to evade real people on the roads in the future.

Reinhard Kleindl

... allow for detection but

prevent identification