DIGITAL CULTURE
ITMO UNIVERSITY

Image and Video Analysis

High School of Digital Culture
ITMO University
dc@itmo.ru

Contents

1 Introduction to Computer Vision **2**

2 Applications of Computer Vision Across Industries **6**
2.1 Medicine 6
2.2 Manufacturing 8
2.3 Video surveillance systems 9
2.4 Biometrics 9
2.5 Image indexing and search 10
2.6 More applications 11

3 Problems and Datasets **12**
3.1 Problems 12
3.2 Datasets 13

1 Introduction to Computer Vision

It is hard to overstate the importance of visual perception for humans. Vision is necessary for almost all tasks in our lives, both personal and professional. It is therefore unsurprising that the problem of computer vision is so important. The ability of a computer to "see" or "analyze" visual information is one of the most common examples of artificial intelligence.

Because humans are so good at effectively gathering and analyzing visual information, we often record and store data about our surrounding world in the form of digital images and videos. Today, there is a camera in every cell phone; cameras are built into home appliances and professional equipment. We can record underwater, from the air, from space. Cameras can be placed inside human bodies or attached to wild animals. The volume of multimedia data increases every second. To get the most out of all this data, we need computers to "see" an image and understand the content. Computer vision is essential in many applications. Here are just some examples where it is used today:

- Diagnosis of brain tumors from the MRI scans (images obtained using magnetic resonance imaging).
- Automatic detection of suspicious objects in surveillance footage.
- A large variety of computer vision algorithms are required for autonomous driving: object detection and recognition are essential for understanding a road scene. Traffic signs, traffic lights, other vehicles, pedestrians, and many other objects on the road need to be detected and recognized.
- Computer vision plays an important role in the world of sports: ball and player tracking, detection of the exact position of the ball, broadcast enchantments, 3D reconstruction of the playfield.
- Biometric systems, for example, unlocking a phone using the owner's face.
- Recognition and translation of signs, posters, restaurant menus from a foreign language.
- Visual quality control systems are one of the earliest and most common applications of computer vision. A typical computer vision task here is to inspect manufactured goods for defects or missing parts.
- Another example is the Amazon Go automated shops, with neither checkout registers nor staff. Upon entering the shop, every shopper registers using an app on their cell phone. Then the cameras track them throughout the

shop and detect items taken by customers from the shelves and placed into their bags.

- The final example on this slide is the application of computer vision in agriculture. Robots and drones in the fields are today's reality. Robots determine the maturity levels of produce, predict its amount, detect diseases, perform automated weeding and harvesting.

These are just a few examples of the possibilities created by computer vision. Later, we will return to a more detailed discussion of these and other applications. Now, let's figure out what computer vision actually is.

Computer vision is a field of science that deals with algorithms and systems to record and analyze images and extract information from digital images and video. In other words, it deals with machines that can see and understand the world as humans do. By analogy with the human visual system, a computer vision system consists of a sensor that records the image and a device that extracts information from the image. Cameras are the eyes of a computer vision system, while the computer and the image analysis and processing algorithms are its brain. Just like the human visual system, a computer vision system receives as input an image and must produce as output some semantic information about the image, some interpretation of the image. For instance, it needs to determine that it's a picture of a garden, that some part of the picture represents a bridge, and some other parts represent trees or flowers. To a human, this appears to be an easy task. But it's not easy to program a computer to do the same.

In this course, we will devote most of our attention to the second, and, in my view, the primary component of a computer vision system: the algorithms and methods of image and video analysis. In literature, different definitions of computer vision can be encountered. Let's present three examples.

- Computing properties of the 3D world from one or more digital images (by Trucco and Veri). In this definition, we suppose that, regardless of whether we exist or not, there is an external world and its visual representation, and we want to understand something about the world by analyzing this representation. This is a definition based on the human visual system. We have a sensor: our eyes; we have a processing device: our brain; and we perceive the world by analyzing the pictures we see.

- Make useful decision about real physical objects and scenes based on the sensed images (by Shapiro). This definition is closer to robotics. We want to make decisions and derive conclusions about real objects around us based on pictures detected by sensors. For example, this definition perfectly fits the description of a robotic vacuum cleaner's function. It decides where to go next and which corner to clean based on what it sees.

- The construction of explicit, meaningful decisions of physical objects from images (by Ballard and Brown). The most general definition of the three. According to it, we simply want to describe phenomena and objects around us using image analysis.

To sum up, we can say that computer vision is about extracting meaningful information from digital images. What is a digital image? There are multiple types of digital images. The most popular ones are raster images. A raster grayscale image is a two-dimensional array of numbers, or a matrix of pixels. Each array element encodes light intensity at the corresponding spatial position. Zero corresponds to the black color, which has zero intensity. How can we determine from an array of numbers that an image is a portrait of Einstein? It's not easy. So, the task of computer vision is to extract meaningful information from a digital image. What meaningful information can be extracted from an image?

- First, it's semantic information. We want to figure out what is shown in the picture, to extract the semantics of the scene. To identify and classify objects, determine their properties and relationships. For example, to determine that a photograph on the left shows a man on a motorcycle with two dogs. To identify the man's identity, the dogs' breeds, the motorcycle's model.

- Second, in addition to semantic information, geometric information about objects is also important. Detection of object sizes, distances between the objects, and their relative positions are important tasks of computer vision. Metric information is needed in robotics and navigation systems, in 3D reconstruction from photos.

The area of computer vision is closely related to many other disciplines. Nowadays, a majority of computer vision tasks are solved with machine learning algorithms. Computer vision is also adjacent to cognitive science, which studies and models cognitive systems; and to neuroscience, which studies neural processes. Computer vision is widely used in robotics: a computer vision system is an essential part of most robots. Information retrieval, speech synthesis and analysis are also close disciplines.

Image processing, image analysis and computer graphics are very closely related areas. Some people consider all three disciplines to be part of computer vision. In this course, we will mostly discuss image processing algorithms. Image processing algorithms take image as an input and produce image as an output. They are used to increase contrast, remove color or noise, apply filters, etc.

In image analysis, it is typical to take an image as an input and produce a certain model or set of features as an output. In other words, given a source image, we calculate certain numeric parameters describing that image. For example, a

gray level histogram, or the coordinates and classes of objects. In image analysis, the result is a set of features. Another related area is computer graphics, where an image is generated based on a model or a set of features.

None of these tasks are possible without knowledge and algorithms from other connected fields, such as pattern recognition and statistics. One can also say that image analysis is a particular case of data analysis and a subfield of artificial intelligence. Another related discipline is neuropsychology: to understand what abilities we have and how image recognition works, it is useful to understand how the brain solves these problems.

The history of computer vision as a research area is considered to begin in the 1960s. A 1996 summer student project in the MIT artificial intelligence group is often brought up as the starting point. This project was the first attempts to detect edges and recover the 3D shapes of simple objects from the "world of toy cubes" (Roberts 1965).

In the 1970s, a basic set of concepts in the area of image processing was formed. Active research was done in edge detection algorithms, first algorithms for stereo matching and optical flow analysis were proposed.

In the 1980s, a lot of attention was paid to the mathematical tools for image analysis. Many papers use Markov random fields to state computer vision problems, which allows us to treat such problems as optimization problems. A theory of levels of image presentation was formed. One of the main landmarks of this period is the book "Vision: A Computational Investigation into the Human Representation and Processing of Visual Information" by David Marr.

In the 1990s, many areas of computer vision continued to develop actively, such as: extraction of object's shape from its motion; methods based on optical flow; 3D reconstruction of objects' shapes from images taken from different view points; image segmentation (normalized cuts and mean shift algorithms). Perhaps the most significant result of this decade was the close interaction of image analysis and computer graphics: mapping a texture from a real picture to a 3D model; construction of a 3D model of an object from images taken from different viewpoints; and automatic generation of other views of the same object. An early example of statistical methods is the use of the principal component analysis for face recognition.

In the 2000s, the connection between computer vision and computer graphics became stronger: the construction of panoramas from multiple shots and the construction of HDR images are examples of applications leveraging methods from both fields. This decade also saw significant shifts in image recognition and classification and information retrieval algorithms. Many methods were proposed for extracting features from images, which were then used as inputs for machine learning algorithms to solve the problems of recognition and classification.

In the last decade, the most significant landmark in computer vision was

undeniably the use of deep learning algorithms. The growth of data volumes and the development of hardware enabled the use of these computationally expensive algorithms, which resulted in significant progress in all computer vision applications over a few years. Before deep learning methods, computer vision systems worked, in the best case, in a controlled environment where lighting, camera, object position, etc. could be controlled. In natural conditions (an uncontrolled environment), there was a huge gap in recognition quality. However, over the last few years, a lot of progress has been made thanks to the use of deep learning methods.

But despite the progress of the last years, we still drive our cars ourselves, and even video surveillance systems often need to be staffed. Why? One of the key problems is the semantic gap.

As human beings, when we look at a picture, we understand its semantics. A computer, on the other hand, understands pixel colors, and with help of modern algorithms can extract the texture and ultimately distinguish a brick wall from a carpet and recognize a person in a photograph, but so far it cannot determine a person's intent based on their facial expression and posture.

Apart from the semantic gap, there are also a series of difficulties that developers of computer vision systems face.

- The same object can appear very different with different lighting;
- An image can be very different depending on the model's pose;
- Background noise and occlusion;
- Overlapping;
- Intra-class variability;
- Angle and location of shooting.

2 Applications of Computer Vision Across Industries

2.1 Medicine

Medicine is one of the primary (and the earliest) application areas of image analysis. Early attempts to automatically analyze medical scans using a computer occurred at the dawn of computer vision. And today computer vision techniques are essential part of medical imaging.

Medical images are an important source of information about the internal organs of the human body. They are actively used for clinical analysis and medical

interventions: diagnosis, planning and performing surgeries, individual modeling of surgical outcomes.

There are numerous ways of creating medical images. Images can be produced using electromagnetic radiation, X-rays, ultrasound, computer tomography (CT) scans, magnetic resonance imaging (MRI). This variety demonstrates how important visual information is in medicine. Most medical equipment used for producing medical images has built-in capabilities for automatic image processing and analysis.

Computer analysis of medical images is applicable in literally all areas of medicine. Detecting fractures on bone X-rays, analysis of chest scans for diagnosing lung diseases (such as pneumonia), analysis of blood vessels using angiography, detecting cancer cells, brain diagnostics, tooth modeling. A wide range of problems in medical image analysis is related to dermatology. The list goes on and on. The slide shows examples of X-ray images.

- A hand X-ray.
- A chest X-ray.
- An aortic angiogram. To obtain an angiogram (an image of blood vessels), a catheter is inserted into an artery or vein. It is threaded into he blood vessel until it reaches the area to be explored. When the catheter reaches the site under investigation, an X-ray contrast medium is injected through the catheter. This enhances contrast of the blood vessels and facilitates detection of any irregularities or blockages.
- A head CT scan.

What kind of algorithms are commonly used for medical image analysis? Can the images be used to detect anomalies, in other words, to tell if this patient's image is different from the image of a healthy person?

Classification algorithms can be used to diagnose diseases. If you have a database of patient scans and you know that the first anomaly occurs in healthy people and the second indicates that the person has cancer, then, using image similarity, you can help doctors diagnose diseases.

Image analysis is also used for individualized human body modeling and to predict treatment outcomes. Although we humans are all similar, every human body has its own individual peculiarities. For example, tooth modeling can be used to show a patient how their teeth will change week to week with the use of braces. Another example is modeling the outcomes of shunting a blood vessel. If a person requires a blood vessel to be shunted, we can determine where to put the shunt by modeling this specific patient's circulatory system using a scan and "inserting" the shunt in this model. This way we'll be able to see how the blood

flow changes and predict how the patient will feel after different scenarios of the surgery.

Image analysis is indispensable for robotic surgeries, non-invasive and minimally invasive surgeries, which are increasingly replacing open surgeries. For example, brain surgeries without opening the skull, such as tumor removal using focused ultrasound. Or various surgeries invoking endoscopic tools and remote control of such tools. In cases of such surgeries, the tools are inserted into the body through the skin or anatomical orifices. The surgery is performed using indirect observation of internal organs through the tools inserted into the body. This way, image processing and analysis algorithms allow surgeons to see and operate.

2.2 Manufacturing

Quality control in manufacturing is another very popular application area of image analysis. Images in visible spectrum (i. e. taken with regular optical cameras) and images obtained from electron microscopes are typically used for this task. Computer vision has long been used for automated quality control of various manufactured goods. All examples on the slide are from a book published in 1992.

- A controller board for a CD-ROM drive. A typical image processing task with products like this is to inspect them for missing parts. The black square on the top, right quadrant of the image is an example of a missing component.

- A pill container. An objective here is to have a machine looking for missing pills.

- Controlling liquid levels in bottles - image processing is used to look for bottles that are not filled up to an acceptable level.

- Controlling plastic quality: this picture shows a clear-plastic part with an unacceptable number of air pockets in it. Detecting anomalies like these is a major theme of industrial inspection that includes other products such as wood and cloth.

- And even quality control of cereal! The next image shows a batch of cereal during inspection for color and the presence of anomalies such as burned flakes.

Early algorithms were based on a large number of heuristics and rules. A new camera location, a new product required a new algorithm. Thus, implementation

of such automated quality control systems was justified only for large manufacturing lines. Today, more and more quality control systems use machine learning and artificial neural networks.

2.3 Video surveillance systems

Another area where image and video analysis algorithms are heavily used is video surveillance. Today, cameras observe us everywhere: in airports, in train stations, in subways, in shops, just on the streets. Given this amount of video data, automated video analysis systems are simply essential. Most cameras are installed to ensure safety. Video recognition systems allow for rapid identification of large groups of people, identify and record facts of their inappropriate behavior (fights, falls, chaotic movement, sudden acceleration, trespassing in a forbidden area), detect forgotten or missing items in the surveilled zone, detect smoke or fire. Suspicious item or behavior detection is not that easy. It is often impossible to provide in advance a description of what should be considered suspicious. It is impossible to imagine all possible examples of suspicious behavior. Therefore, such systems often model normal behavior and detect anomalous behavior by identifying deviations from the norm.

Apart from security purposes, video surveillance systems are often used to study human behavior. For example, cameras installed in a shop can be used to detect which sections of the shop are more popular and use this information to arrange goods more conveniently.

Video surveillance systems enable automatic collection of fees for driving on a toll road, issuance of speeding tickets, observation of the animal world. The list of applications is unlimited!

2.4 Biometrics

Another area which actively uses computer vision is biometric identification systems. These systems use people's intrinsic physical characteristics to verify their identification. Sometimes these characteristics are called a person's biological code.

An advantage of biometric identification systems compared to traditional ones (for example, PIN-based or password-based systems) is that it identifies not some external object belonging to the person, but rather the person themself. The characteristics being analyzed are inseparably linked to the person; they cannot be lost, transferred, or forgotten, and they are extremely difficult to forge. In addition, these characteristics are not subjected to wear and do not require replacement or repair.

Among primary biometric technologies that use computer vision are identification based on fingerprints, face geometry, hand geometry, eye iris or retina, vein pattern geometry, signature, and others.

Early biometric systems were rather expensive and required significant computational resources. Today we unlock computers and phones using fingerprints or face geometry.

Let me present another story related to biometrics, which has made a lot of noise in the media. There is even a Wikipedia article about this story. In 1980s, a National Geographic photographer took a photo of an unidentified Afghan girl. The photo became very famous after it appeared on the cover of National Geographic magazine in June 1985. Because the girl's identity was unknown, the photograph was called simply "Afghan Girl". The photo is sometimes compared to Leonardo da Vinci's Mona Lisa portrait and is called "Afghan Mona Lisa". Over the course of the years 1990–2000 the journalist made several attempts to learn the girl's name. Several women claimed themselves to be the "Afghan girl"; many young men claimed that the "Afghan girl" is their wife. 17 years after the photo was taken, the photographer found her in Afghanistan and confirmed her identity using biometrics, which demonstrated a full match between her iris and the one depicted in the photograph on 1980s.

2.5 Image indexing and search

Image indexing and search comprise another important problem which relies on image analysis algorithms. There are different kinds of image collections and archives.

- Personal. For example, a person can make a couple thousand photos during a vacation, which need to be handled in some way afterwards.
- Professional. They consist of millions of photographs. There is also a need here to somehow organize them, search them, find something that is necessary.
- Collections of replicas. These also consist of millions of images. Nowadays, numerous museums have virtual presence, with collections of digitized replicas. One interesting image retrieval problem for this type of images is to find all paintings by the same artist. A person can determine, based on style, that they're looking at, let's say, a painting by Salvador Dali. It would be nice if a machine could do that too.

How to arrange all these collections? What other tasks can be solved? One can construct a system for navigating the collections by classifying them by topics. Put bears in one class, elephants in another, oranges in the third, to make it easier to navigate this collection.

Another task is duplicate detection. In two thousand vacation photos, there are way fewer non-repeating ones. We like to experiment, vary exposure, focal

distance, etc., which gives us a large number of "fuzzy" duplicates, i. e. approximate copies. In addition, duplicate search can help detect an unlawful use of your photograph that you might have one day uploaded to the internet.

Another interesting problem is the selection of the best frame. An algorithm can help determine which picture the user likes the most. For example, if it is a portrait, the face should be lit, the eyes open, the picture should be sharp, etc. Modern cameras do have such a function.

Another search problem is the creation of collages, i. e. selection of pictures that will look well next to each other.

2.6 More applications

There are many more practical problems where image analysis is used.

- Tracking and aiming systems in the military industry. Obvious examples are detecting enemy soldiers and vehicles and rocket control. The most perfect rocket control systems send a rocket to a prescribed area rather than at a specific target, and the target is selected when the rocket reaches the area, based on the video data it receives.

- Autonomous vehicles, both for military and general use. The degree of autonomy varies from fully autonomous (unmanned) to vehicles with computer-vision-based systems that assist the driver or the pilot in various situations. Fully autonomous vehicles use computer vision for navigation, i. e. for gathering information about their location, for mapping their environment, for obstacle detection. Examples of such systems are obstacle warning systems in cars and autonomous landing systems in planes. Systems for autonomous driving are being actively developed and tested today.

- Computer vision is also actively used in the entertainment and film industries to create special effects and to merge real actors' performance with computer graphics. Today, filmmakers can record the facial expressions of a real actor and superimpose them on a drawn character or on the face of another actor. Computer vision is now an inseparable part of an absolute majority of video games, virtual and augmented reality applications.

- Another major area of visual processing is remote sensing, which usually includes several bands in the visual and infrared regions of the spectrum. Images of population centers are used routinely (over time) to assess population growth and pollution. Satellite images are used to automatically map a region. Weather observation and forecasting are other important applications of satellite imaging. Hurricane detection and prediction, wildfire detection, and even changes in ground level as indicators of groundwater depletion are all done by analyzing satellite images of Earth.

- Digitization of printed documents, text detection and recognition are additional applications of image analysis algorithms. Today there are apps that can translate foreign-language signs and notices in real time.
- And, of course, computer vision is an inseparable component of robot navigation and control systems.

3 Problems and Datasets

3.1 Problems

We've discussed a number of practical applications of computer vision. So how can a computer be taught to "understand" images for all these applications to work? Let's watch a video. Your goal is to count how many times the players in white pass the ball. Did you notice that a person dressed as a bear passed by the players? Let's watch the video again. After I told you about the bear, you obviously noticed it. Because you know what to look for.

This video demonstrates that even the human visual system usually solves only one problem at a time, ignoring signals that are irrelevant to the problem. Likewise, when designing a computer vision system, people usually solve only one problem. This is much easier that trying the solve the abstract problem of "teaching a computer to understand the image". What are the most common problems in various computer vision systems?

- Face detection: find all human faces in a picture.
- Face recognition: determine which person a face corresponds to.
- Face identification: determine whether a given face corresponds to a specific person.
- Posture recognition: determine which posture from a predetermined set a given human posture corresponds.
- Identifying areas that are homogeneous by color or texture.
- Object recognition: determine which one of a predetermined set of classes a given object in a picture corresponds to.
- Semantic segmentation: determining the areas of an image pertaining to one object.

- Image classification: determining which category an image or the objects depicted in it belong, without localizing the objects in the image. Classification can be used to answer such questions as whether the image contains an elephant, or a plane, or a sunset, without specifying where exactly the elephant or the plane is in the picture.

- Text detection and recognition: determine where in a picture text is located and what it says.

Even a task as general as semantic description of an image can be solved as a combination of more narrowly defined problems. For example, let's consider this photo. If I ask you to describe this photo, I will most likely hear that this photo shows a little girl eating an ice cream. You can describe what she's wearing, where she is.

It is possible to construct a similar description of an image automatically using a computer by solving problems like object and face recognition, determining a person's sex and age, identification of areas of homogeneous color, action recognition, texture extraction.

3.2 Datasets

Publicly available datasets collected for validation and comparison of various algorithms play a very important role for the development of the field of computer vision (as well as many other branches of data analysis). Standard datasets are necessary for objective comparison of different approaches to solving the same problem. Data is essential to develop machine-learning-based algorithms.

Let's familiarize ourselves with the most well-known datasets used by the scientific community to develop and test more and more perfect image analysis algorithms.

Lena. Perhaps the most famous test image used for testing and demonstrating various image processing algorithms is "Lena". The test image is a digitized portrait of a Swedish model, which in turn is a fragment of a Playboy magazine centerfold. According to Wikipedia, in 1973, Alexander Sawchuk from the University of Southern California needed a photo portrait with good dynamic range for illustrating an article about image processing. Sawchuk scanned a fragment of a Playboy poster. He used a scanner with a 100 dots per inch resolution, producing a 512 by 512 points image. Soon this picture turned into a de facto industry standard: it was used to test and develop all sorts of image correction mechanisms and to master new processing algorithms.

MNIST. One of the first image sets for testing machine learning algorithms is the MNIST collection (short for "Modified National Institute of Standards and Technology"). It consists of samples of handwritten digits. This dataset was used to develop the first neural-network-based algorithms for image analysis. The

MNIST collection contains 60,000 training images and 10,000 test images of size 2828 pixels. All images are divided into 10 classes, one for each digit. The task of recognizing handwritten digits in this dataset is the classical classification problem. Given an image, it is necessary to determine which class it belongs to (i. e. which of 10 digits is depicted). By today's measures, both the collection size and the image resolution are rather small. But this dataset is still actively used, mostly as a first example in deep learning and for simple tests with neural networks. Programming a neural network for training and testing using this dataset is like writing a "Hello World" program in a new programming language.

CIFAR. CIFAR-10 and CIFAR-100 are also very popular datasets commonly used in machine learning and computer vision research. They are similar to MNIST by image count and resolution, but the images are in color and more diverse. CIFAR-10 dataset contains 60,000 32x32 color images in 10 different classes, such as planes, cars, birds, and others. CIFAR-100 dataset has 100 classes containing 600 images each. These datasets are also widely used to test image classification algorithms.

ImageNet. ImageNet is a labeled image collection, whose creation is considered by many to be the beginning of the deep learning era and one of the main reasons of incredible progress in computer vision with deep learning. Training deep neural networks requires a lot of data. ImageNet is the first large dataset of annotated images. ImageNet contains over 14 million images (compared to 60,000 in MNIST and CIFAR!), divided into more than 20,000 classes. Since 2010, the creators of ImageNet run an annual competition called ILSVRC (ImageNet Large Scale Visual Recognition Challenge). In this challenge various software products compete in classifying and recognizing objects and scenes in the ImageNet database. ILSVRC uses a subset of the ImageNet collection consisting of about 1.2 million images and 1,000 classes. Images are 224x224 pixels.

PASCAL VOC. Before ImageNet and ILSVRC, the most well-known image classification and object detection competitions was PASCAL VOC challenge (Pattern Analysis, Statistical Modelling and Computational Learning Visual Object Classes), conducted from 2005 to 2012. Image collections gathered for this challenge were for a long time a de facto standard for testing various algorithms, especially object detection and segmentation algorithms. PASCAL VOC, unlike previously discussed collections, contains object-level markup, where the specific regions containing objects are marked in each image. `http://host.robots.ox.ac.uk/pascal/VOC/databases.html` `http://host.robots.ox.ac.uk/pascal/VOC/images/tud3c.html`

Caltech Caltech101 and Caltech256 are two other image collections commonly used to test image classification and object recognition algorithms. These datasets consist of 101 and 256 image categories, respectively. The size of each image is approximately 200300 pixels, just like in Im-

ageNet and PASCAL VOC. Most Caltech101 categories contain about 50 images, but there are also categories with many more images. For example, the "plane" and "face" categories contain about 800 samples. Each image is annotated with the coordinates of a box containing the object and with the object class. http://www.vision.caltech.edu/Image_Datasets/Caltech101/ http://www.vision.caltech.edu/Image_Datasets/Caltech256/

MS COCO All previously considered collections contain images of relatively small sizes. However, most cameras today produce images at higher resolution. And for a number of problems (for example, object detection for autonomous vehicles) it's important to work with high resolution images. MS COCO (Microsoft COCO: Common Objects in Context) contains images with the resolution of up to 1000 pixels along the longer side. Also, most collections that we've discussed previously do not have markup for semantic segmentation, where every image's pixel is marked as belonging to a particular object. The MS COCO collection is considered today to be a standard for comparing object detection and semantic segmentation algorithms. It contains 328,000 images with objects from 91 categories. The total number of marked objects is about 2,500,000. http://cocodataset.org/ https://arxiv.org/pdf/1405.0312.pdf

Cityscapes. Another widely used high-resolution image set is Cityscapes. The collection contains images of cityscapes collected in 50 cities in Europe. Each image is 10242048 pixels. The dataset contains 5,000 images with detailed object annotations, on the level of individual pixels, and about 20,000 images with less detailed annotations. https://www.cityscapes-dataset.com/ https://arxiv.org/pdf/1604.01685.pdf

Labeled Faces in the Wild. Apart from databases with images of objects of various categories, there are also more specialized datasets, some with images belonging to a single category. There are annotaed datasets for pedestrians detection (http://coding-guru.com/popular-pedestrian-detection-datasets/), gesture recognition, road sign recognition.

Special attention has always been given to the problem of face detection and recognition. A large number of annotated collections have been created for these purposes. One of the most widely used collections today is Labeled Faces in the Wild: http://vis-www.cs.umass.edu/lfw/.

As it follows from the name, the collection is designed for studying the problem of unconstrained face recognition. It contains images of faces collected from the web, without any constraints to lighting, posture, or head position. It's not a collection of portraits with a monotonous background, but rather still frames from news reports. The collection contains over 13,000 images of famous people, found throughout the internet. About 1,680 people from the collection have more than one photograph.

Medical images Despite the fact that analysis of medical images is one

of the most significant practical applications of computer vision, there aren't many publicly available collections of medical images. This is mainly due to the private nature of such data. Only in recent years, when it became clear that the availability of large datasets can significantly speed up the development of algorithms, did large anonymized collections of medicals scans begin to appear. The first such collection that gained popularity within machine learning groups was ChestXray14, containing about 112,000 chest scans with indicators of 14 kinds of diseases. Recently a Stanford machine learning group released another large collection, CheXpert. This collection contains 224,316 chest scans. `http://academictorrents.com/details/557481faacd824c83fbf57dcf7b6da9383b3235a` `https://stanfordmlgroup.github.io/competitions/chexpert/` `https://arxiv.org/abs/1901.07031`