ISSN 1870-4069

Designing New CAPTCHA Models Based on the Cognitive Abilities of Artificial Agents

Edgar D. García-Serrano, Salvador Godoy-Calderón, Edgardo M. Felipe-Riverón

Centro de Investigación en Computación, Instituto Politécnico Nacional, CDMX, Mexico

edserrano18@outlook.com, sgodoyc@cic.ipn.mx, edgardo@cic.ipn.mx

Abstract. A CAPTCHA is a program that allows or denies access to services by generating and grading test that humans can pass but current computer programs cannot. Captchas are generally used to protect free web services from automated programs. Also, they can provide an idea of which fields in artificial intelligence are the most explored and which could be researched in the near future. Most of the tests that are based on text recognition have been broken by optical character recognition (OCR) techniques while those based on images are vulnerable to machine learning attacks. Humans make use of cognitive abilities to carry out tasks in daily life, even if they are not conscious which ones or how they use them. The current state of technology is still not enough to reproduce some human cognitive abilities, and the identification of those abilities is the basis for the design of new CAPTCHA models. In this paper we introduce seven new models of CAPTCHA to test some cognitive abilities that are supposed to be beyond the capabilities of artificial agents. We use some APIs to proof that images used in the proposals are extremely difficult to be recognized by artificial agents.

Keywords: CAPTCHA, human interaction proofs, Turing test, cognitive abilities.

1 Introduction

CAPTCHA is acronym for Completely Automated Turing Test to Tell Computers and Humans Apart [1], term that was introduced first time in the year 2000.

A CAPTCHA is a program that automatically generates a task or test that most humans can solve but artificial agents cannot. Captchas are based on a Turing test [2], which is a question-answer game with a human judge and two participants; a human and a computer program. The judge does not know which of the participants the computer is, and by means of a question series the judge has to identify the computer while both participants try to convince the judge that they are humans. In captchas, the judge is a computer program, and instead of a question series, it generates a task. If the task is completed successfully, the user is considered to be human, otherwise

the user is considered an artificial agent. If a computer program correctly solves a CAPTCHA, it is considered that the CAPTCHA model has been compromised or *broken*. Some uses for these tests are: online poll protection, web services registration protection, dictionary attacks prevention, blog spam comment avoidance, and many more.

Generally, captchas are used as a security measure known as question-answer authentication, and its algorithms are totally public. For this reason, breaking a CAPTCHA is an interesting artificial intelligence problem. In other words, it would be difficult to make a program able to solve CAPTCHA even if it is known how it works exactly. According to [1], a CAPTCHA is a win-win situation. On one side, if the CAPTCHA is not broken, then we keep a way to differentiate humans from computers. On the other side, if the CAPTCHA is broken, then a useful AI problem is solved.

Since its creation, captchas have motivated multiple researches in AI fields, from OCR techniques as in [3], breaking classic text captchas or machine learning with deep learning methods [4] to break semantic image captchas. In this research we left out the security context and we use the CAPTCHA as an opportunity for some cognitive computing researches. In this way, we consider some principles and features of the set human cognitive abilities for designing seven new captcha models. The tests present in these captchas are based on capabilities present in humans and which computers cannot emulate yet. Furthermore, since these are new captcha models there is no an artificial agent trained to solve them. Instead, we separately tested Google, Microsoft and IBM's agents with the abstract images included in the proposed captcha models.

2 Related Work

Given the wide variety of captchas currently around internet, it is possible to set a taxonomy [5, 6] in which we can identify two main groups: visual captchas and audio captchas.

Within the visual captchas class, we can find text-based captchas which are characterized by displaying a set of characters with some transformations and distortions; the user must input the characters that he can recognize. Text-based captchas are vulnerable to OCR and some examples are the classic CAPTCHA and ReCAPTCHA models [7]. Other kind of visual CAPTCHA are those based on images; it is generally a test in which the user must identify some animals or objects and select them correctly according to the instructions. An example of this is the ASIRRA captcha model [8] that is a simple image categorization test. Another captcha model is based on videos, known as video-captchas, where the task is to recognize and input the red characters, part of the banner that crosses the screen in some direction. The NuCaptcha model [9] was a pioneer in these tests, although it is vulnerable to more sophisticated OCR. The next captcha models are mathematical in nature, this is to say, the user must solve a basic arithmetic problem to get access. Finally, in the puzzle-captcha, like its name suggest, it is necessary to complete a simple puzzle to complete the task. The last two

types of captchas are often difficult for humans, but they are even more difficult for artificial agents.

Audio captchas, as an alternative for the existing captchas models, strictly require only some eye-sighting abilities to pass the test. Visually handicapped persons generally fail the test and so they are denied the corresponding services that the CAPTCHA was guarding. For this reason, it seems necessary to draw on other elements to design captchas models, for example, sound. Audio captchas make use of sound to present the instructions and the task to be solved. They consist in a fragment of audio in which a person pronounces letters or numbers and the user must type the symbols listened.

So far, webservers only use two of the CAPTCHA types above, most common are the text-based ones (already broken by OCR techniques) and image based (broken by deep learning algorithms). As they have been broken, a way of make them stronger is searched. However, this leads to a more complex task even for humans. So this search for stronger captcha models tend to create new tests to be implemented and that require the use of different types of cognitive abilities, for which artificial agents are faced with bigger challenges.

Emotion identification requires an intense use of certain cognitive abilities, for that reason a 2-layer CAPTCHA was proposed [10] where the first layer consists in a puzzle of an image describing an emotion. The second layer asks the user to select from a menu, the word that better describes the presented image. The cognitive abilities required for solving this type of CAPTCHA are mainly visual processing.

To support the fact that human cognition plays an important role while solving captchas, in [11] an experiment developed which faces many users with image based captchas in varying difficulty levels. The author observes that a captcha modeling design process that adapts models based on some cognitive factors could effectively improve results.

In [12] the human ability of humor understanding is used; according to the author, humor understanding is the most advanced cognitive ability of a human being. He proposes a captcha model where the user is presented with four images that represent different stages of some funny story, and the user has to sort these images in the proper order.

There are some researches where the design of new tasks in captcha models is strongly based on psychological principles. Such is the case of [13], where constructivist theory and mental models are the base for such design. The main consideration is that our brains do not perceive images pixel by pixel, instead, they are built with models that sum up what the senses perceive (Gestalt laws). Captchas proposed in this work are based on Gestalt laws and Geon's theory of pattern recognition. Memory and context play an important role in how humans visually interpret objects.

The captcha models proposed in our research are based on tasks that are still difficult for artificial agents to solve. For example: abstraction ability, common sense reasoning, visual perception (Gestalt laws), and attention. Background knowledge also plays an important role in the development of what we call Cognitive Captchas (C-Captchas).

3 Proposed Captcha Models

CAPTCHA models herein proposed are based in human cognition, conceived as the set of all processes by which sensory input is collected, transformed, reduced, elaborated, stored and used [14]. To identify exactly all of those processes is a difficult task in which even psychologists have not come to an agreement. Following is the list of human cognitive abilities required to solve these tasks in those captcha models:

- Attention.
- Short-term memory retention.
- Visual/spatial processing.
- Natural language understanding.
- Fine grained motor skills.
- Executive functions.
- Common sense reasoning.

Even when there is not a formal definition of cognitive abilities for artificial agents, certain techniques or algorithms exist that emulate some human abilities. Here we list the artificial agents' abilities that correspondingly take part in the knowledge acquirement:

- Pattern recognition
- Natural Language processing.
- Graphical interface interaction.
- Machine Learning.
- Associative binding.

We focus on those abilities where the artificial agents face greater problems than humans. State of the art techniques in digital image processing are insufficient for adequately emulating human perception abilities. Pattern recognition and machine learning algorithms suitable for solving this kind of tasks usually require large amounts of input data. For that reason, we selected a set of images that we identify as abstract, because of their lack of textures, colors and closed outlines. All these images show objects or characters in an indirect way, either with cartoons or suggested by blurred strokes. Humans are usually able to recognize these images even with insufficient or simplified information thanks to a combination of perception, visual processing, and background knowledge processing. We refer to this combination of abilities as abstraction.

Another ability useful during the design process was common sense reasoning. It is well known that artificial agents possess strong logical reasoning but not common sense. Common sense is a set of knowledge that is acquired through experiences. Research in this kind of reasoning has been carried out [15] but it is complex and difficult to formalize; it is necessary to study the simplest cases to focus on the problems that we want to solve.

A set of minimum features is established for all captchas to keep a certain level of difficulty. These features are:

- Every CAPTCHA has three graphical regions: instructions, challenge and solution.
- Instructions in natural language must present a common sense challenge.
- All images must be abstract.
- The user has just one chance to solve the CAPTCHA.

3.1 Model 1: Story Completion

This model contains two challenges. Firstly, the natural language challenge where a simple instruction is provided to complete a story that entangles a random word representing an emotion (keyword). There are three rows (see Fig. 1) with two abstract images and an empty frame; every row represents a different story. The goal is to relate the keyword from natural language challenge with one of the three stories, select the ending frame from the solution region and complete the sequence by dragging the chosen image to the empty frame of the corresponding story. This constitutes the common sense reasoning challenge.

3.2 Model 2: Object-environment Association

The instructions in this captcha model simply tell the user to make proper associations in each case but do not refer to specific objects nor they give more information. In the challenge region three abstract images of animals are presented (drawn by line patterns). In the solution region several abstract images representing places or environments can be found. The user is asked to use markers with specific shape to relate each animal with its usual environment (see Fig. 1). Abstraction and background knowledge processing are required in order to solve this captcha model.

3.3 Model 3: Differentiating Feature Identification

This captcha model is defined as a semantic classification challenge according to the features of several animals presented in nine independent and abstract images (see Fig. 1). All but one of the animal images share a specific feature and the remaining image does not. The user is asked to select the image that does not share the common feature and drag it to the solution region option that represents the feature in question. In this case abstraction, natural language processing, pattern recognition and background knowledge processing are required.

3.4 Model 4: Foreground and Background Image Composition

The goal is to identify the foreground and background parts of a provided image, select them among the images in the challenge region and drag them into the solution region. The conceptual composition of the selected foreground element and background element images should result in the provided images (see Fig. 1). The images that represent the foreground and background elements in the challenge region are not

exactly the same as the one provided even though the image conceptually. A high level of abstraction is required for the resolution of this captcha model.

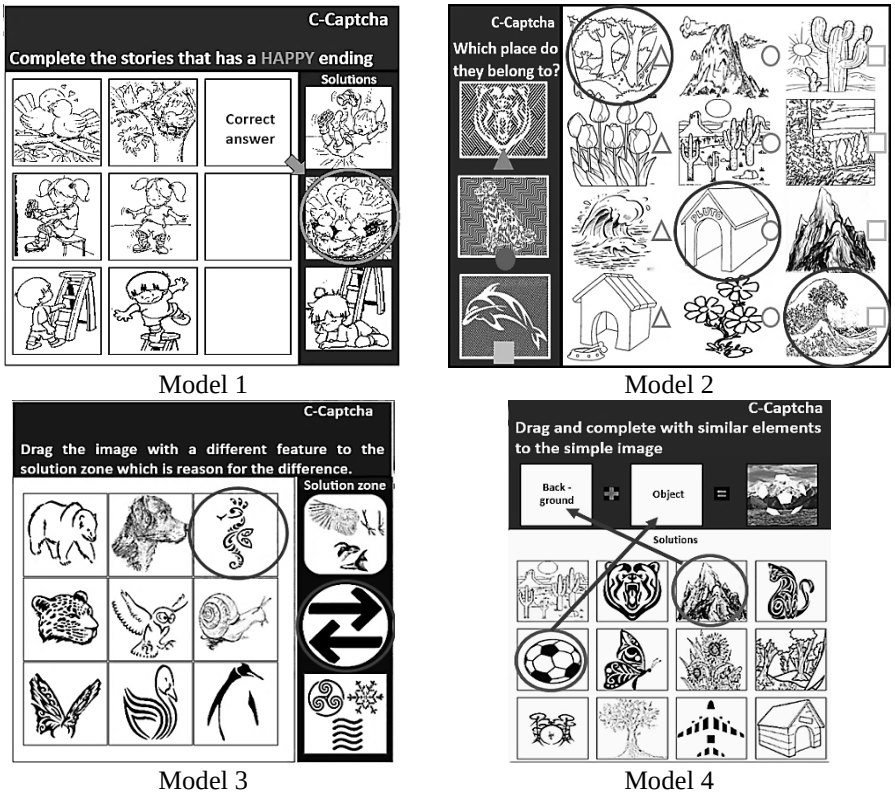

Fig. 1. Captchas from model 1 to model 4 showing the correct solutions and the correct places.

3.5 Model 5: Shared Features Identification

Abstraction is tested in this captcha model, a simple task in which the user needs to drag and drop the answers that match with the feature mentioned in the instructions region (see Fig. 2). The challenge region presents twelve images; from which the user has to select three images whose associated entities share the common feature. Natural language processing, abstraction, common sense reasoning and background knowledge processing are required for this challenge.

3.6 Model 6: Size Based Sorting

The goal is to sort, from smallest to biggest, four abstract images representing different sized objects (see Fig. 2). After recognizing each abstract image, the size sorting can only be performed by associating each image semantics with background knowledge about its represented object size.

3.7 Model 7: Suggested and Explicit Object Relation

The abilities needed to complete this CAPTCHA are: pattern recognition, common sense reasoning and abstraction. It is a simple test in which the user must drag the abstract image from the challenge region that corresponds with the suggested object shown in the solution region (see Fig. 2).

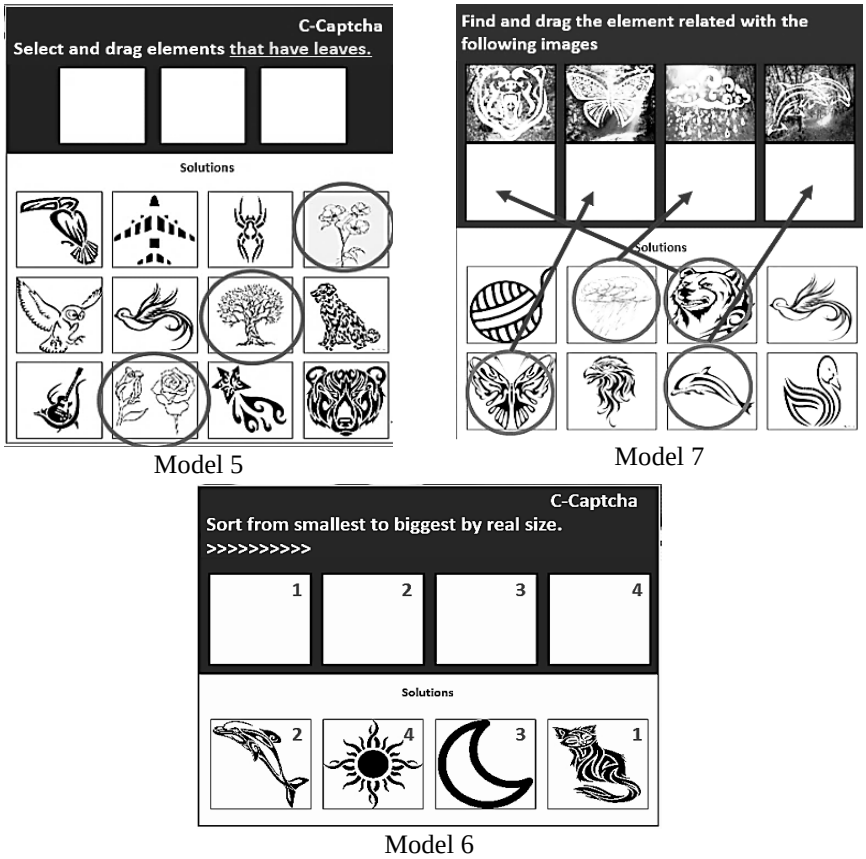

Fig. 2. Captcha models 5 to 7 showing the correct solutions and their corresponding places.

4 Testing Artificial Agents

Since proposed captcha models are new there are no current artificial agents trained to solve them. Therefore, the only possible test is to present the abstract images for recognition to specialized cognitive agents. In order to make a comparison between humans and artificial agents, some experiments were carried out using four types of images: photography, realistic drawings, sketches and abstract images. The selected agents were: Microsoft's Computer Vision [16], IBM's Visual Recognition API [17],

and Google's Vision API [18]. Each agent is trained to input and image and processes it to provide some information as the content, colors, size and semantics.

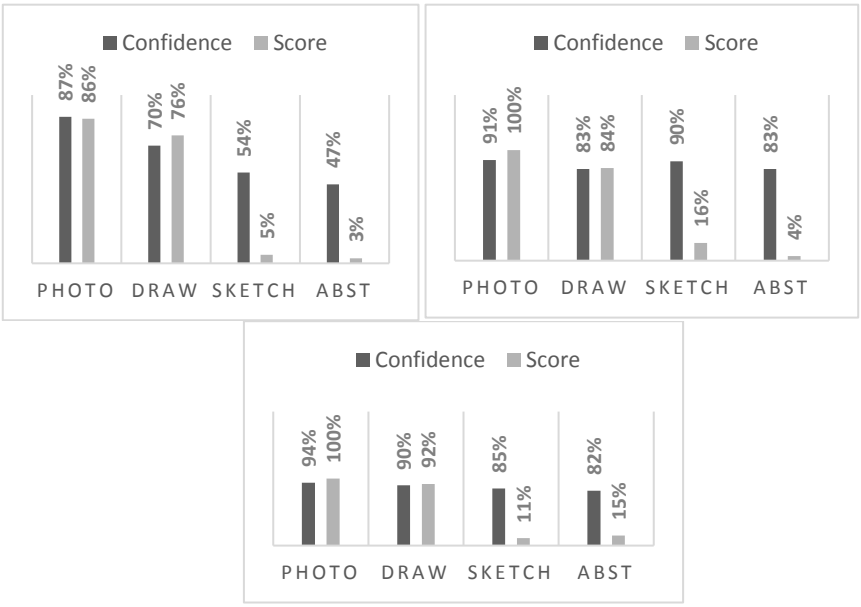

Fig. 3. Results of image recognition test: Microsoft 's API (up left), IBM's API (up right) and Google's API (below).

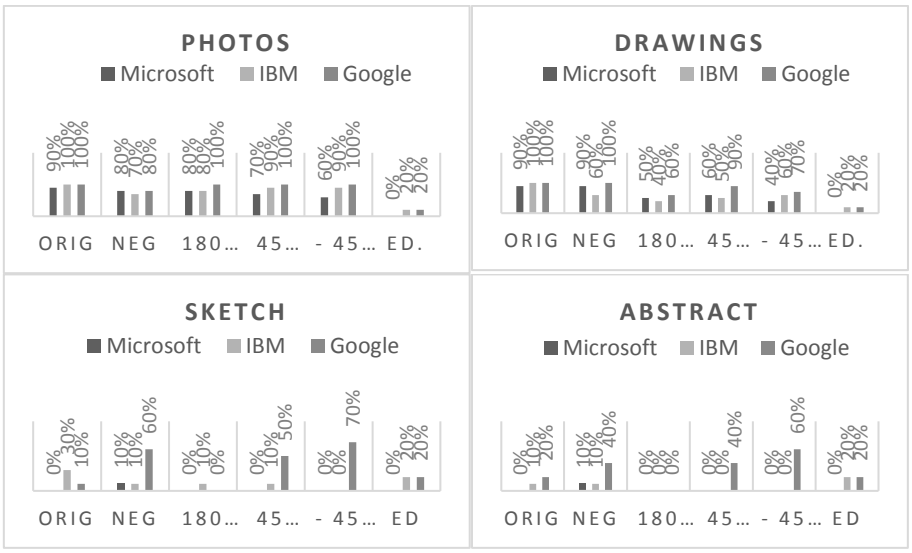

Fig. 4. Score comparison for each kind of images and its transformations

The first experiment used a corpus of 400 images, 100 images of every kind of image above mentioned. The results in image recognition for each agent are shown in

Fig. 3. The graphics show the number of correct classifications and the confidence level; the latter is a mark that each agent provides assessing its own classification. All agents can classify the photos and drawings with high confidence, but as the abstraction level increases the score gets reduced. When sketches and abstract images as those used in C-Captchas are presented to the selected agents, the score gets lower even when they show a high confidence.

In the second test, we took 10 images of each type and four transformations were applied to them. The transformations were: inverted colors and rotations to 45 °, -45° and 180°. Also, 10 photos were modified with Photoshop to make them look like abstract images; the resulting images were added to the experiment. The agents were tested with a total of a 170 images. The results were compared among them to support the idea that abstractions and transformations make the task more difficult in case of artificial agents (see Fig 4).

As it was expected, the transformed images were more difficult to recognize that the original ones. In the case of abstract images, transformed images were almost never correctly classified by agents. Strangely Google 's agent better classifies sketches and abstract images if they are in negative colors or with rotations of 45°. This phenomenon could probably be explained with the image that Google 's team used as the training set for the agent. It should be mentioned that Google 's agent had the best performance when classifying the images, but it is still low with abstractions.

So we conclude that abstract images used in the proposed captcha models represent great challenges to artificial agents as they have low rate of success in this experiment. If an artificial agent tries to break this captchas, first it must understand the instructions, recognize the images, and finally devise a solution to the test. Now, an artificial agent that has low rate of success recognizing abstract images is expected to have even a lower rate of success passing the whole test.

5 Conclusions

Advances in artificial intelligence have allowed traditional captchas to be broken, motivating some research to improve currently used captcha techniques. To introduce more sophisticated aspects of some cognitive abilities into captchas. Those abilities include common sense reasoning, abstraction, and visual processing as explained by Gestalt's laws. For the moment it seems not possible to replicate those cognitive abilities with artificial agents.

Abstract images, by themselves, represent a big problem for artificial agents, even for those specialized in image recognition. Enterprises like Google, Microsoft or IBM cannot handle high levels of abstraction yet.

If abstract images are added to a captcha model based on common sense and requiring natural language processing or having a specific given context, the difficulty in recognizing images increases. A future goal is to determine which areas of AI could constitute new sources for future solutions, in order to create new captchas principles that be easy for human beings and complex for artificial agents.

References

1. Von Ahn, L., Blum, M., Langford, J.: Telling humans and computers apart automatically. Communications of the ACM, 47(2), 56–60 (2004)
2. Turing, A. M.: Computing machinery and intelligence. Mind, 59(236), 433–460 (1950)
3. Baecher, P., B üscher, N., Fischlin, M., Milde, B.: Breaking reCAPTCHA: a holistic approach via shape recognition. Future challenges in security and privacy for academia and industry, 56–67 (2011)
4. Sivakorn, S., Polakis, I., Keromytis, A. D.: I am robot: (deep) learning to break semantic image captchas. In: Security and Privacy (EuroS&P), 2016 IEEE European Symposium, pp. 388-403. IEEE (2016)
5. Sheheryar, M. A., Mishra, P. K., Sahoo, A. K.: A review on Captcha generation and evaluation techniques. ARPN Journal of Engineering and Applied Sciences, 11(9), 5800–5811 (2006)
6. Kaur, K., Behal, S.: Captcha and Its Techniques: A Review. International Journal of Computer Science and Information Technologies, 5(5), 6341–6344 (2014)
7. CAPTCHA Homepage, http://www.captcha.net, last accessed 2017/06/25.
8. Elson, J., Douceur, J. R., Howell, J., Saul, J.: Asirra: a CAPTCHA that exploits interest-aligned manual image categorization. In: ACM Conference on Computer and Communications Security. (Vol. 7), 366–374 (2007)
9. Nu Captcha Homepage, http://www.nucaptcha.com, last accessed 2016/04/18.
10. Tanvee, M. M., Nayeem, M. T., Rafee, M. M. H.: Move & select: 2-layer CAPTCHA based on cognitive psychology for securing web services. International Journal of Video & Image Processing and Network Security, 11(5), 9–17 (2011)
11. Belk, M., Germanakos, P., Fidas, C., Holzinger, A., Samaras, G.: Towards the personalization of CAPTCHA mechanisms based on individual differences in cognitive processing. In: Human Factors in Computing and Informatics, 409–426 (2013)
12. Yamamoto, T., Suzuki, T., & Nishigaki, M.: A proposal of four-panel cartoon CAPTCHA. In: Advanced Information Networking and Applications, 2011 IEEE International Conference, 159–166. IEEE (2011)
13. Rusu, A., Docimo, R.: Securing the web using human perception and visual object interpretation. In: Information Visualisation, 13th International Conference, 613–618. IEEE (2009)
14. Neisser, U.: Cognitive psychology: Classic Edition. Psychology Press, New York (2014)
15. Davis, E., Morgenstern, L.: Introduction: Progress in formal commonsense reasoning. Artificial Intelligence, 153(1–2), 1–12 (2004)
16. Computer Vision API Homepage, https://azure.microsoft.com/es-mx/services/cognitive-services/computer-vision/, last accessed 2017/05/17.
17. IBM Watson Developer Cloud, https://visual-recognition-demo.mybluemix.net/, last accessed 2017/05/22.
18. Google Cloud Platform, https://cloud.google.com/vision/, last accessed 2017/05/22.