Quick, Stat!: A Statistical Analysis of the Quick, Draw! Dataset

Raul Fernandez-Fernandez (✉), Juan G. Victores, David Estevez, and Carlos Balaguer

All of the authors are members of the Robotics Lab research group within the Department of Systems Engineering and Automation, Universidad Carlos III de Madrid (UC3M).
rauferna@ing.uc3m.es

Abstract. The Quick, Draw! Dataset is a Google dataset with a collection of 50 million drawings, divided in 345 categories, collected from the users of the game Quick, Draw![1]. In contrast with most of the existing image datasets, in the Quick, Draw! Dataset, drawings are stored as time series of pencil positions instead of a bitmap matrix composed by pixels. This aspect makes this dataset the largest doodle dataset available at the time. The Quick, Draw! Dataset is presented as a great opportunity to researchers for developing and studying machine learning techniques. Due to the size of this dataset and the nature of its source, there is a scarce of information about the quality of the drawings contained. In this paper a statistical analysis of three of the classes contained in the Quick, Draw! Dataset is depicted: *mountain*, *book* and *whale*. The goal is to give to the reader a first impression of the data collected in this dataset. For the analysis of the quality of the drawings a Classification Neural Network was trained to obtain a classification score. Using this classification score and the parameters provided by the dataset, a statistical analysis of the quality and nature of the drawings contained in this dataset is provided.

Keywords: Quick, Draw! Dataset, Statistical Analysis, Neural Networks.

1 Introduction

Since the introduction of Deep Learning techniques there has been a tendency to increase the number of layers in order to increase the performance of these architectures [1,2,3]. As a result of this tendency, in the present, it is common to find Deep Neural Network architectures with hundreds of layers. These architectures require very large amounts of training data in order to obtain useful results. To address this problem, some techniques have been proposed to deal with the problem of lacking sufficient amounts of training data [4] by generating new data from the limited available data. However, having a rich and diverse dataset is still a much more effective way to correctly train a Neural Network (NN) [5,6,7]. The Quick, Draw! Dataset [8] is presented as the largest sketch dataset available at the time. The Quick, Draw! Dataset collects the sketches of the players of the Quick, Draw! open online game. In this game, the players are given a category they have to sketch in a maximum time limit of 20 seconds. At the same time, a classification Neural

[1] https://quickdraw.withgoogle.com/

Network attempts to guess the category that the player is drawing. The player wins if the NN correctly guesses the category of what the player was prompted to sketch within the time limit.

What is more interesting about this dataset is that it does not store the sketch performed by the player as a bitmap matrix composed by pixels. Instead, the sketch is saved as time series composed by pencil positions (x,y). Therefore, in addition to the spatial information that can be contained in an image, this dataset also contains temporal information of the sketch. As additional information, with each sketch, the dataset also include the category the player was asked to sketch, the timestamp when the sketch was created, the country where the player was located, and a unique sketch identifier.

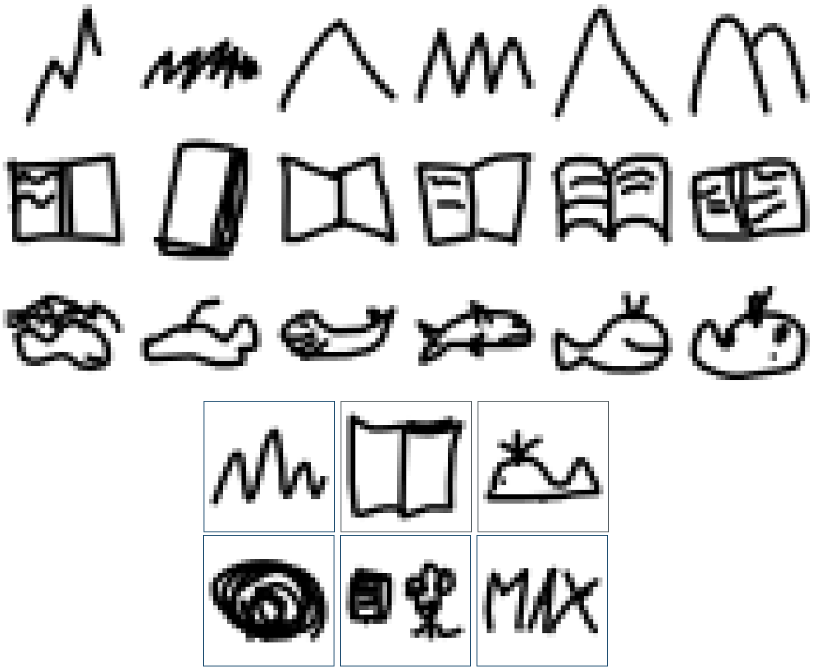

Fig. 1. First three rows: Random examples extracted from the Quick, Draw! Dataset of the three categories studied in this paper. From top to bottom: *mountain*, *book*, *whale*. Last two rows: The first row depicts good sketches with high classification scores. The second row depicts examples of bad sketches with low classification scores. From left to right: *mountain*, *book*, *whale*.

In this paper, a statistical analysis of three of the categories presented in this dataset: *mountain*, *book* and *whale*, is performed. These three categories were selected to be different enough in terms of complexity to be considered valid representatives of the wide spectrum of sketches in the Quick, Draw! Dataset. The *mountain* category is considered an example of a low complexity sketch; the *whale* category is a representative of a more complex spectrum of sketches; while the *book* category can be considered an average complexity sketch, see Fig. 1 for reference. From each of the sketches presented in each of the categories, three parameters are extracted: the classification score of the sketch, extracted using the classification Neural Network; the number of strokes performed by the player; and the sketch length, defined as the total number of line segments that defines the sketch. We consider the rest of the available information (country, timestamp...) less relevant for the study performed in this paper.

For the classification Neural Network, a Deep Convolutional Recurrent Neural Network with Long Short Time memory (LSTM) layers, explained in Section 4, is implemented. This NN is trained using the Quick, Draw! Dataset. After training, each of the sketches is assigned a score. This score is the ouput assigned by the classification Neural Network to the sketch category, given the sketch as an input to the network. The rest of the parameters can be directly obtained from the dataset.

The remainder of this paper is organised as follows. First, a study of related works using the Quick, Draw! Dataset is presented in Section 2. Section 3 depicts a more detailed introduction to the dataset. Section 4 introduces the architecture of the Classification Neural Network implemented. The setup used for the experiments is explained in Section 5. To finish, the final section of results in Section 6 and conclusions in Section 7.

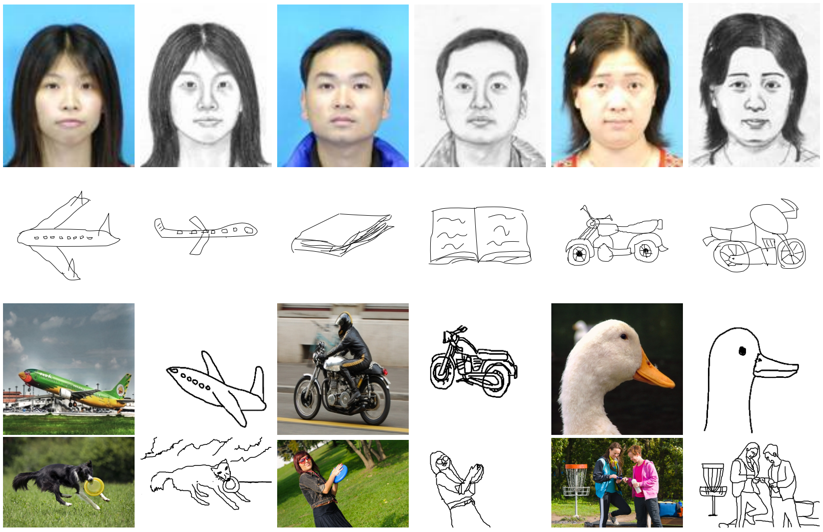

Fig. 2. Examples extracted from alternative datasets to the Quick, Draw! Dataset. Each row depicts random examples extracted from a different dataset. From top to bottom: CUHK Face Sketch Database [9]; Eitz dataset of human sketches [10]; the Sketchy Database [11] and Li countour dataset [12].

2 State of the Art: Sketch Datasets

In addition to the Quick, Draw! Dataset, there are a number of alternatives where the authors created sketch datasets in order to train their machine learning models. The CUHK Face Sketch Database contains a total of 606 sketches of human faces, each with their corresponding photo [9]. In the original paper, the authors proposed a Markov Random Field model to synthesize photos to sketches and vice versa. Eitz et al. proposed a dataset containing 20 thousand sketches divided in 250 categories [10]. Two different machine learning methods, Nearest-Neighbor and Support Vector Machines, were used for the classification of the sketch dataset. The results of these methods were compared with the ones obtained with human classification. The Sketchy database [11] presented by Sangkloy et al. is a collection of more than 70 thousand sketches, each paired with a photo, divided in 125 categories. Here, the authors used Convolutional Neural Networks to map the photo and sketches to the same feature space. Recently, Li et al. presented the contour drawing dataset [12]. The countour drawing dataset contain 5000 sketches with 1000 paired photos that can contain multiple objects. The original application of this dataset was to train a deep learning method for sketch generation. All of these datasets are open and can be downloaded in the web page of each project.

3 Quick, Draw! Dataset

The Quick, Draw! Dataset contains a total of 50 million sketches divided in 345 categories. This dataset is the result of extracting the sketches provided by the players of the online game Quick, draw!. Compared to state of the art alternatives, it is much larger and may contain much richer information. As such, the quality of the sketches may experiment important variations, which may be positive for training and posterior recognition.

Sketches are stored in the ".ndjson" format[2]. Each sketch is a structure divided in different fields containing the following information: a unique id; the category where the sketch belongs; a recognition flag indicating if the sketch was recognized by the classification network; the timestamp of the sketch; the country of the player; and an array with the drawing information. This drawing information consists of N time series, where N is the number of complete strokes. Each complete stroke is encoded as a $3 \cdot T$ matrix, where the first two rows correspond to the "x" and "y" position of the pencil, the third one "t" with the timestamp, and "T" is the number of positions contained within the stroke.

As an alternative to the raw dataset, a simplified dataset is proposed by the Quick, Draw! Dataset authors. In this dataset, the authors introduce the following simplifications:

- Align the sketch to the top-left corner, and assign a zero value to that position.
- Scale the sketch to a 255x255 canvas.
- Re-sample the sketch to have a 1 pixel spacing within strokes.

[2] http://ndjson.org/

- Use the Ramer–Douglas–Peucker algorithm with an epsilon value of 2.0 to simplify the strokes.
- Delete the timestamp information.

This simplified dataset is available in three different formats: ".ndjson", ".bin" and ".npy". In this paper the ".ndjson" version of the simplified dataset is used as the input to the classification Neural Network. The ".npy" version is used for the visualization of sketches.

4 Neural Network Classification Model

The classification Neural Network takes the sketch input as a sequence of pencil positions of "x" and "y" and infers the category where the sketch belongs. Due to the nature of the dataset, a Recurrent Neural Network (RNN) using LSTM layers was implemented in this paper [3]. The idea is that in addition to the spatial information contained by the pencil position, there is also a latent space containing relevant temporal information for defining the category of the sketch. This is where RNN excels.

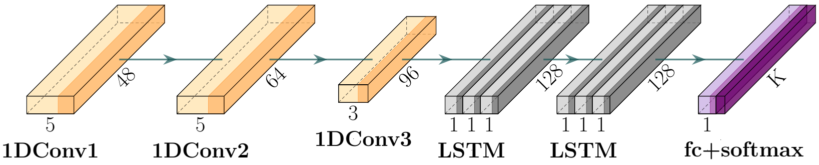

Fig. 3. Neural Network classification model.

The classification Neural Network model used in this paper is depicted in Fig. 3. The input of the NN is first passed through three 1D convolution layers. These layers relate the information of the input that is temporally close. The ouptut of the convolution stage is passed to two LSTM blocks using a *tanh* activation function. The LSTM blocks are composed by three LSTM layers each. The final layer is a Fully Connected (FC) layer with a softmax activation, being the outputs the score assigned for each of the K categories, where K is the number of training categories. For the training of the Neural Network, a total of 10000 sketches were used for the training stage with 1000 sketches for the validation stage for each category.

5 Experimental Setup

The goal of this paper is to provide a statistical analysis of the Quick, Draw! Dataset. Three parameters will be extracted to perform this analysis: the score obtained with

[3] Adapted from: https://tinyurl.com/tensorflow-models-quickdraw

the classification Neural Network proposed in section 4; the total sketch length; and the total number of complete strokes defined in each sketch. The Quick, Draw! Dataset is composed by 345 categories, each of these containing hundreds of thousands of sketches. Training a NN that is able to classify all of the categories is computationally expensive. In this paper we choose three of these categories to train the classification Neural Network. The goal is to obtain representative statistical results that can give the reader a first impression of the spectrum of sketches contained in this dataset.

The three categories studied in this paper are: *mountain*, *book* and *whale*. These three categories go from the simplicity of a mountain to the complexity of a whale, being the book the average level sketch example. The three parameters used in this paper were chosen to extract the most relevant information from the sketches. The classification score provides a value of the quality of the sketch. The sketch length and number of strokes provides information about the complexity of the sketches. Together, these three parameters provide information about the quality and complexity of the sketches.

In the following section, the results of these experiments are depicted. All the sketches contained in each of the categories studied in this paper at the time of its writing were used to extract these results.

6 Experimental Results

At the time of writing this paper, a total of 364.409 sketches were contained over the three categories studied. The distribution of these sketches between the categories was the following: 128.541 *mountain* sketches; 119.365 *book* sketches and 116.503 *whale* sketches. All of the sketches were used in order to obtain the results presented in this paper.

Table 1. Experimental results for the statistical analysis of the Quick, Draw! Dataset as a function of the three categories studied in this paper, where μ is the arithmetic mean, and σ is the standard deviation.

	Classification score		Number of strokes		Sketch length	
Category	μ	σ	μ	σ	μ	σ
mountain	5.16	2.13	2.09	2.15	18.04	15.08
book	2.96	2.67	7.21	3.78	42.50	17.17
whale	4.44	2.49	5.55	3.45	45.38	17.68

Table 1 depicts the arithmetic mean and the standard deviation of the distribution of sketches as a function of the three parameters studied in this paper. From these results, it can be concluded that the *book* is the category with a lower average classification score. This is interesting, since the *whale* category was initially a more complex one, while it has a higher average classification score. There are two possibilities that can explain this behavior: either people perform poorly at drawing

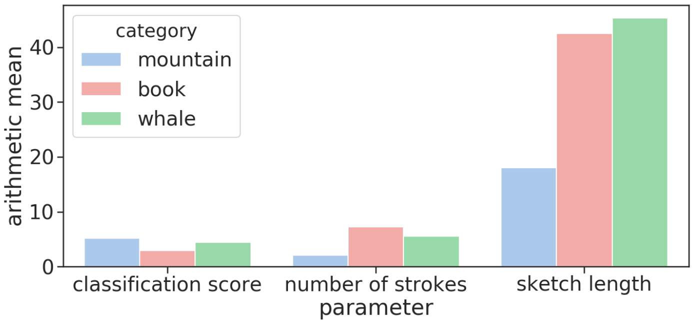

Fig. 4. Comparison between the results obtained with the three categories studied in this paper.

books or the classification network performs poorly at book classification. If we take a look at the second and third parameters depicted on the table, in average, *book* sketches are composed by a 30% more strokes than *whale* ones do with a similar length in the total drawing. Although the concept of drawing a book looks easier than that of drawing a whale, people tend to draw more complex books than whales. This can lead to a loss in the quality of the drawings, or a loss in the classification network performance. Fig. 4 provides a graphical comparison between the results obtained for each of the three parameters studied as a function of the category.

Table 2. Number of sketches as a function of its z-score divided by categories. Only the classification score is taken into account to compute this z-score.

	z-score (classification score)									
category	$-\sigma$	σ	-2σ	2σ	-3σ	3σ	-4σ	4σ	-5σ	5σ
mountain	3042	2034	1725	273	498	12	60	7	3	4
book	10245	2687	1383	126	267	16	69	2	21	0
whale	2064	5165	78	279	15	51	9	11	0	7

Table 2 contains the number of sketches as a function of its z-score. This z-score gives an impression of the number of outliers in the dataset. In the case of the *mountain* and *book* sketches the majority of outliers are in the negative region. In these two categories the population of outliers are mainly composed by people performing poor sketches. In the case of the *whale* category, the population of outliers is larger in the positive region meaning an outlier population of people performing better sketches. This gives us a first impression of the differences in the complexity of the sketches. In the two first categories, the more outlying sketches are the ones that are poorly drawn. In the case of the *whale* the more outlying sketches are the one that achieve the higher level of classification score. The results in this table also show how the number of outliers is not as high as expected for

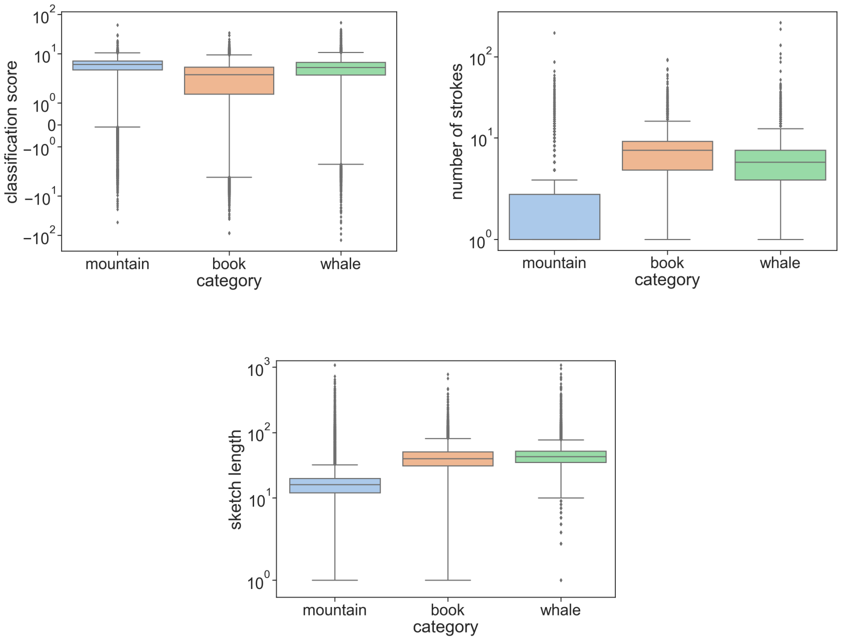

Fig. 5. Comparison between the results obtained with the three categories studied in this paper.

a dataset of this nature. In the case of the *mountain* category, around 96% of the population of mountain sketches are contained in the range $(-\sigma, \sigma)$, 89% for *book* sketches, and 94% for *whale*. Here, it should be noted that, as extracted from the Quick, Draw! Dataset repository, the dataset authors claim to have moderated the collection of drawings individually. Fig. 5 depicts various boxplot graphs for a better visualization of the distribution of sketches as a function of the three parameters and the category where they belong.

Fig. 6 depicts the relation between the classification score and the number of strokes and length of the sketches. The first thing to take into account is that both the sketch length and the number of strokes have a similar effect in the classification score achieved in each category. In the case of the *mountain* category, the peak of classification score is achieved with low values of these two parameters. Increasing the number of strokes or the sketch length comes with a loss in the classification score for this category. More simple sketches are preferred by the neural network for the classification of *mountain* sketches. The case of the *book* category in terms of shape is quite similar. The only difference here is that this peak is achieved with higher levels of strokes and sketch length. As in the *mountain* category, once reached this peak, the classification score of the sketches decrease. In the case of the *whale* category, increasing the complexity of the sketches comes in average

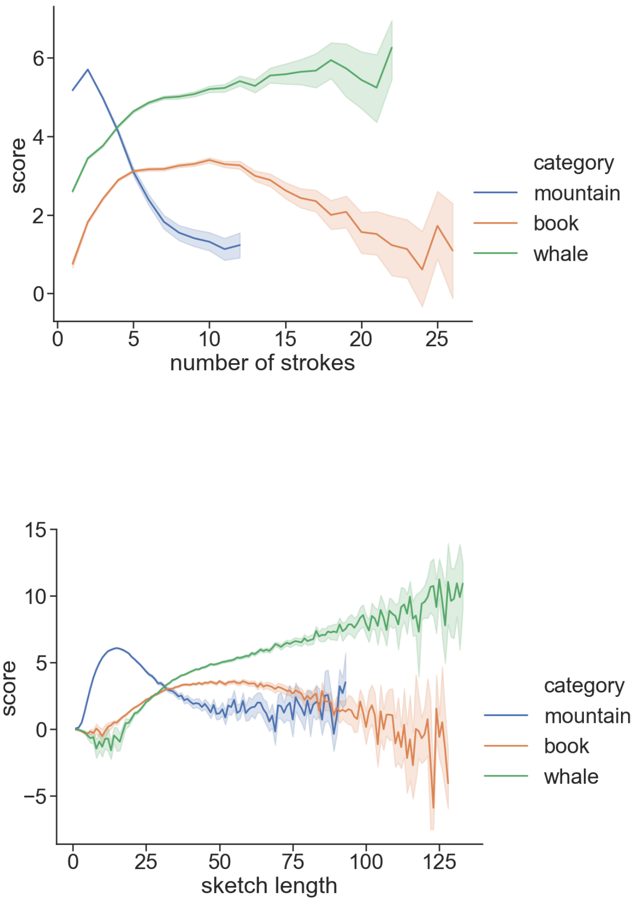

Fig. 6. Classification score as a function of the sketch length and number of strokes for the three categories presented in this paper. Sketches with a z-score bigger than 5 were not depicted in the graph for clearness reasons.

with an increase of the classification score obtained by these sketches. In this case, high complexity sketches are, in average, better in quality. This comes with a cost: higher complexity sketches also mean higher levels of variability in the quality of sketches. This effect can be found in the three categories studied; however, it is more relevant in the case of the *whale* category, since the best sketches are the ones with a higher complexity. A similar effect happens for low quality sketches in the case of the *book* and *whale* categories. For these two categories low quality sketches suffer from high variances in quality. This effect does not occur in the case of the *mountain* category.

7 Conclusions

The current tendency to increase the number of layers in Deep Neural Networks have produced models with hundreds of layers. These models require very large amounts of data to provide useful results. The Quick, Draw! Dataset, composed by over 50 million sketches, is presented as way to deal with this need of massive data for training. In this paper, a statistical study of three of the categories present in this dataset: *mountain*, *book* and *whale* is presented. The results depict the distribution of all the sketches of these three categories of the dataset as a function of three parameters: classification score, number of strokes, and sketch length. For the experiments, a classification Neural Network using LSTM layers was trained and used for prediction. The results obtained in this paper have depicted some interesting results regarding the nature of the sketches contained in this dataset.

The first relevant conclusion is that the complexity of a sketch is not only limited to the complexity of the category as would be expected. Conceptually, describing a *whale* is more complex than describing a *book*, however the results obtained in this paper show that in average, people drew more complex sketches to define a *book* than to define a *whale*. A similar concept was already introduced by [11], where a “sketchability” criterion was defined as a subjective measure of how hard it would be to sketch some of the photos. Some interesting studies can be extracted about the differences in sketch complexity between the different categories.

Another interesting result extracted from the experiments is that the number of outliers presented in the dataset is surprisingly low for a dataset of this nature. The Quick, Draw! Dataset’s source is the open online game Quick, Draw!, where hundred of thousands of players around the world have accessed this game and sketched. With this premise, one would expect for this dataset to have high variances in quality between the sketches in a category, see Fig. 1 for reference. The results in the experiments depict that more than 90% of the sketches in average are contained in the range $(-\sigma, \sigma)$ for the categories presented. In addition to this, the number of outliers outside the range $(-3\sigma, 3\sigma)$ is lower than the 0.05% of the total population in each category studied. In the repository of this dataset, the authors claim to have performed a individual analysis of the sketches. This could explain the results obtained. However, there is no information about how this analysis was performed, and as Fig. 1 depicts it was not perfect. This should be taken in account before considering this dataset a raw dataset or considering all the sketches in this dataset raw examples of each category.

8 Applications and Future Work

The Quick, Draw! Dataset may fit in two kinds of applications: applications where very large amounts of data are required (e.g. training Deep Neural Networks) and applications where the trajectory performed during the sketch, rather than the image, is required. Certain precautions must be taken into account during the interpretation of its contents, as stated in the dataset repository[4] and depicted in the results of this paper, this is not a raw dataset. Any conclusions extracted from the analysis of this dataset could be altered by possible exclusions resulting of the reviewing process. Future works using this dataset involves a more complete analysis using more categories as input, and the study of more parameters that can be interesting for machine learning techniques.

For downloading, obtaining more information or checking interesting projects involving this dataset, refer to the online repository referenced in this paper.

9 Acknowledgment

The research leading to these results has received funding from RoboCity2030-DIH-CM, Madrid Robotics Digital Innovation Hub, S2018/NMT-4331, funded by “Programas de Actividades I+D en la Comunidad de Madrid” and cofunded by Structural Funds of the EU.

References

1. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11211 LNCS, 294–310. https://doi.org/10.1007/978-3-030-01234-2_18
2. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 7-12-NaN-2015, 1–9. https://doi.org/10.1109/CVPR.2015.7298594
3. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
4. Perez, L., & Wang, J. (2017). The Effectiveness of Data Augmentation in Image Classification using Deep Learning. arXiv:1712.04621
5. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[4] https://github.com/googlecreativelab/quickdraw-dataset

6. Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv:1708.07747.
7. Krizhevsky, A., & Hinton, G. (2009). Learning multiple layers of features from tiny images (Vol. 1, No. 4, p. 7). Technical report, University of Toronto.
8. Jongejan J., Rowley H., Kawashima T., Kim J., & Fox-Gieg N. (2016). The Quick, Draw! - A.I. Experiment. https://quickdraw.withgoogle.com/
9. Wang, X., & Tang, X. (2009). Face Photo-Sketch Synthesis and Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(11), 1955–1967. https://doi.org/10.1109/TPAMI.2008.222
10. Eitz, M., Hays, J., & Alexa, M. (2012). How Do Humans Sketch Objects?. ACM Trans. Graph. (Proc. SIGGRAPH)
11. Sangkloy, P., Burnell, N., Ham, C., & Hays, J. (2016). The Sketchy Database: Learning to Retrieve Badly Drawn Bunnies. ACM Trans. Graph. (Proc. SIGGRAPH)
12. Li, M., Lin, Z., Mech, R., Yumer, E., & Ramanan, D. (2019, January). Photo-Sketching: Inferring Contour Drawings from Images. In 2019 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 1403-1412). IEEE.