

A Novel Algorithm to Classify Hand Drawn Sketches with Respect to Content Quality

Ochilbek Rakhmanov(✉)

Nile University of Nigeria, Plot 681, Cadastral Zone C-OO,
Research and Institution Area, Jabi, Abuja, Nigeria
ochilbek.rakhmanov@nileuniversity.edu.ng

Abstract. In this paper, the methodology of a novel algorithm called Counting Key-Points algorithm (CKP) is presented. The algorithm can be used during classification of same type of hand drawn sketches, where content quality is important. In brief, the algorithm uses reference pictures set to form vocabulary of key points (with descriptors) and counting how many times those key points appeared on other images, to decide the image content quality. CKP was tested on Draw-a-Person test images, drawn by primary school students, and reached 65% of classification accuracy. The results of the experiment show that the method is applicable and can be improved with further researches. The classification accuracy of CKP was compared to other state-of-art hand drawn image classification methods, to show superiority of the algorithm. As the dataset needs further studies to improve the prediction accuracy, it would be released to the community.

Keywords: Classification · Sketches · Bag of visual words · Machine learning · Draw-a-Person test

1 Introduction

Sketch can be defined as a freehand drawing, which can be done by professionals, amateurs, adults and even children. With the development of touch screen digital devices like tablets and phones, sketch research has become an active and popular field in computer vision. Many users will sketch shapes on electronic device screens, which can create tremendous amount of the data for further researches. For instance, once of such recently released datasets contains more than 50 million free-hand sketches [1], which was collected during a mobile game Quick Draw by Google. The front-runner in this field was the research and the dataset released by Eitz et al in 2012, with 20,000 sketches [2].

Aforementioned publicly accessible datasets triggered many researchers to conduct machine learning classification experiments on hand-drawn sketches [2–4]. Unlike the ordinary photographic images, hand-drawn sketches differ in three main aspects: (i) they can be very abstract and deformed and still represent the same object, (ii) some people may draw all features of the object others may miss them, and lastly, (iii) sketches lack colors as they are usually black and white only [5].

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12252, pp. 179–193, 2020.
https://doi.org/10.1007/978-3-030-58811-3_13

The drawings of the children have high importance on their mental development, educational performance and can reflect inner world of the child [6]. For this purpose, researchers have developed many different types of cognitive tests, which require a child to draw a shape of particular object, and later can be used for evaluation of various mental and physical conditions [6–9]. Thus, hand-drawn sketches have high importance in the fields of education and healthcare. Cognitive drawing tests proved to be effective tool for adults with mental disorders as well [9–11]. Many different types of drawing test were developed, and several studies been done on their reliability and validity. As this research addresses the issue of classification of hand drawn sketches with respect to content quality, one particular cognitive test were chosen in this context, DAPT – Draw-a-Person test [7].

All of the datasets released to community [1, 2, 12] have one thing in common, they are useful to classify different objects; cup from plate, lion from horse, cat from dog etc. A computational device should be trained to find difference between objects to classify them. But this is not same if one decides to train computational device using the images collected from DAPT, because in this case the important feature for classification are not the differences, but common features and how frequently they are appearing in the image. In other words, the content quality of the image is important in the classification procedure.

1.1 Content Quality

At this point, a new field should be addressed, the content quality of hand-drawn images. Content quality might be defined as content that delivers value, that solves a problem. In this, if an image contains more of main features of the object in comparison to other image of the same object, then it should be classified as higher class. For instance, Fig. 1 shows 3 different level of 2 different objects, face and bicycle. If they need to be classified with respect to image quality, definitely the content of the image would carry high importance. Type C is better drawn in comparison to Type B, while Type B is better than Type A, thus, they need to be classified accordingly.

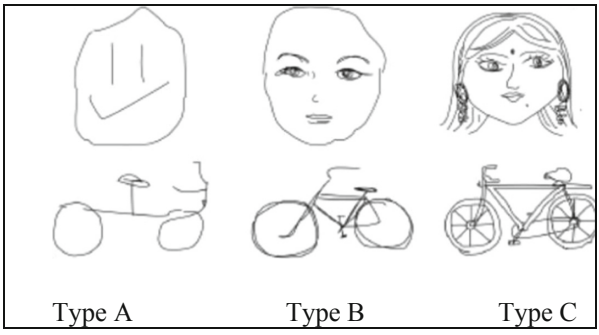

Fig. 1. Different representation of the face and bicycle [4]

Just like in Fig. 1, content quality is becoming main actor in classification of images, collected during DAPT procedure. Figure 2 present some sample images from DAPT test, drawn by children which belong to 3 different age categories. From Fig. 2, it is clearly evident that content quality of images was main actor in manual classification of the pictures to different classes (Fig. 3).

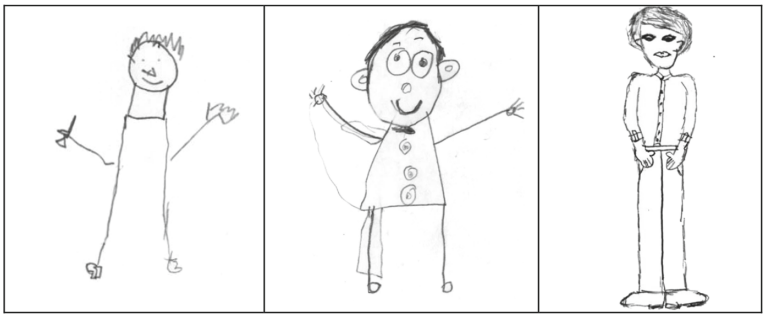

Fig. 2. DAPT sketches which belong to different categories [13].

2 Background of the Study

2.1 Machine Learning and Computer Vision Methods for Hand Drawn Sketch Classification

K-means, KNN, SVM [14] and CNN [15] are mostly used machine learning algorithms for classification of hand-drawn sketches, while computer vision algorithms like SIFT-Scale-Invariant Feature Transform [16], HOG – Histogram of Oriented Gradients [17] and BoVW – Bag of Visual Words [18] are extremely helpful for researches during future extraction, alongside with commonly used image processing operations like contour detection and convolution of the image [19].

In 2012, Eitz et al collected 20,000 original sketches of 250 different objects [2] and released this dataset to community, before testing it for classification with HOG and SVM, to reach 56% of accuracy. Year later, Li et al used several techniques like Ensemble matching, Star graph with KNN to improve classification accuracy to 62% [20]. Next promising research was conducted by Schneider et al in 2014, where they employed SIFT and SVM alongside with Fisher vectors [3] to reach accuracy of 67%. Several progressive researches were conducted by the same group Yu et al in 2015 and 2017 [4, 21], and CNN was successfully tested for Eitz et al dataset to reach 74% and 77% classification accuracy. CNN seemed to be an acceptable tool, but in 2016, Ballester and Araujo trained two state-of-art CNN structures, AlexNet and GoogleNet, and observed that those CNN structures are resulting with unacceptable accuracies, below 50% [22], when they were trained with hand-drawn sketches from Eitz et al dataset. This issue needs a further clarification in future researches. In one of recent studies, Rakhmanov tested the aforementioned classification algorithms (SVM, ANN)

without computer vision algorithms (SIFT, HOG,BoVW), just using pixel values, on dataset provided by Eitz et al [5]. He stated that existing methods' results are reasonable to accept.

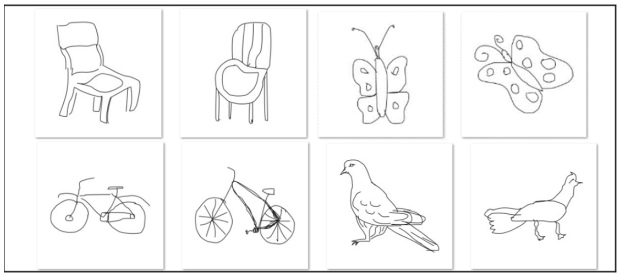

Fig. 3. Sample pictures from Eitz et al dataset (chair, butterfly, bicycle and pigeon) [6].

In 2017 a research group from Google [1] released a dataset of hand-drawn sketches, which is presently appears to be world largest doodling dataset, with more than 50 million samples. They used RNN to produce a software which completes the sketch from initial strokes. This triggered more researches; Xu et al [23] and Bensalah et al [24] used GNN (Graph Neural Network) in combination with CNN to classify the images. Still the highest classification accuracy is remaining in region of 77%.

Table 1 presents researches and used techniques in chronological order. While Eitz et al dataset was employed mostly, future researches may concentrate to Ha et al dataset, as it is a relatively new dataset.

Table 1. Chronological development table

Used technique	Publication	Dataset used	Accuracy
HOG + SVM	[2]	[2]	56%
KNN+Star graph+ SVM	[20]	[2]	62%
Fisher + SIFT + SVM	[3]	[2]	67%
CNN+ Stroke removal	[21]	[2]	74%
CNN + Data Augm.	[4]	[2]	77%
GNN + CNN	[24]	[1]	77%

2.1.1 Observations

Following observations can be done from Table 1:

a) If deep learning is employed to classify hand-drawn sketches, then it should be combined with some different techniques, to avoid biased results, as Balester and Araujo stated that implementing CNN straight, may result with unexpected low accuracy.

b) If the technique used is not of deep learning type, SVM can serve as good classifier.
c) BoVW uses HOG during methodology, and HOG uses SIFT. It appears that BoVW – HOG – SIFT can be also employed as reliable tools to classify hand-drawn sketches, and results can be promising.

2.2 Draw-a-Person Test (DAPT)

First conceived by Dr. Florence Goodenough in 1926 [7], the Draw-a-Person test is a skill test that is designed to measure a child's mental age through a figure drawing task. It estimates the progress of learning visual, cognitive, and motor skills by having the candidate draw a human figure, scoring the drawing for presence and quality of figure features, and comparing the score to children's typical rate of acquisition of figure features [7]. Throughout the years, this test underwent many discussions, whereby the researchers tested the validity and reliability of the test, usually resulting in supportive conclusions. The instrument is among the top 10 tools used by practitioners, according to Yama et al [25]. It is widely used in early childhood education; primary school counselors can use it to monitor children's mental development. In Psychology, for instance, it has been used during comparison between healthy patients and those with mental disorders. There are many supportive researches and case studies for these aforementioned fields, summarized by Naglieri et al [26]. Due to the wide adoption and use of DAPT, there was a need to conduct some experiments on DAPT sketches, to see if we can develop a model to classify images automatically. Evaluation of DAPT can be done only by professional psychologist or practitioner. But there are very few schools in low income countries (including both private and public) who have a psychologist as working personal, because it is too expensive for them. Thus, development of such automatic tool can help to millions of schools, guidance-counsellors and teachers around the world.

2.2.1 How the DAPT Scored?

This section will provide a brief explanation of the DAPT picture (as it was proposed by Goodenough [7]). In the DAPT image evaluation, we count the number of features, like eyes, nose, mouth, hair, hands, shoulder, fingers, etc. Yet, some more scores come from the comparison of the geometrical positioning of the features. For instance, a child can draw one leg short and other long or provide a sketch of eyes whose proportion on the face is abnormal. The sum of these scores results in a total score (minimum 0, maximum 51 points). We look up for corresponding mental age from score $\sim$ mental age scale proposed by Goodenough [7]. This mental age is compared to biological age, and if the difference is high, then counselor or practitioner advices that the child needs special assistance or training on recognizing the functions of the objects surrounding him or her in his or her daily life [7]. Figure 4 shows some pictures from our dataset and their corresponding mental ages.

Figure 5 demonstrates some human figures which belong to the same class, even though they are drawn in very different ways. Thus, the biggest challenge in this dataset as it can be observed from Fig. 5 is that sketches may appear in very different shapes. Unlike in most common image classification tasks, we are not actually looking for the

difference between the images, but for the common features and how frequently they appear. The exploration of any possible solution for this challenge is the main goal of this research.

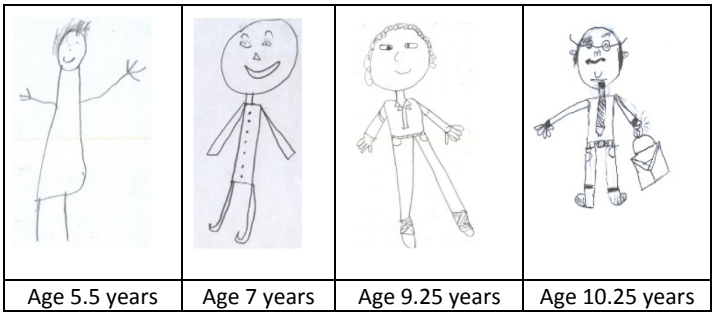

Fig. 4. Sample sketches of the students.

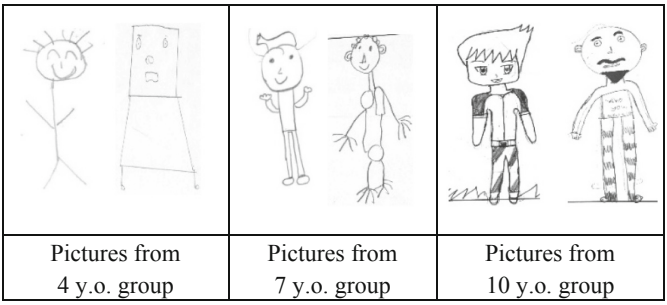

Fig. 5. Different human shapes from same group.

Another important difference is that the Eitz et al dataset was collected through touch screen devices and mostly participants were adults. All pictures are clear and sharp. However, it was impossible to do so in our case as we had kids aged 4–7 years who were unfamiliar with touch screens or other electronic devices. Thus, our dataset consists of sketches drawn on a plain sheet. Each student tried to draw his or her best picture; so, there are many pictures where students tried to erase previous sketches and replace them with new ones. They also tried to draw extra elements and the lines in their drawings were clashing or not straight. All this created significant noise on images which affected prediction accuracy in some manner.

As from computer science aspect, it is very important to mention that a child's picture can be only in one highest class. In other words, a picture can only have highest class, like picture of 5 years old child, or picture of 9 years old child, with respect to content quality. This is advantage of the scoring system of DAMT, it cannot be misclassified between 4 years old and 8 years old, or 5 and 10. We only must identify what is the highest class the image belongs, which is making this challenging task

easier to approach. In other words, we don't have to predict exact class of the picture, we need to predict what is the highest class it can belong.

3 Purpose of the Study

The main objective of this study is to present the methodology on how to develop and apply CKP algorithm for image classification of hand drawn sketches with respect to their content quality and to present comparative results with respect to other existing methodologies to prove the superiority of the CKP algorithm, as this type of sketches require extra-ordinary approach. A unique dataset of DAPT images (with more than 1000 images) was collected in order to conduct this experiment.

An early preliminary version of this work was published by Rakhmanov et al [27], where initial results of the experimentations on DAPT images were published. However, this paper presents complete guide on how the methodology of the CKP works, and presents a complete comparative study on results.

4 Instruments

We used 2 important concepts in machine learning and computer vision to establish CKP algorithm, they are K-means and BoVW. We also compared the prediction accuracy of CKP to other widely used algorithms. In this context, we compared CKP results with SVM+HOG sketch classification method [2], and with CNN [4]. We described the complete development process of CKP in Sect. 5, while SVM+HOG and CNN were only briefly described in methodology section, as they are already explained in the literature.

- *K-means*. K-means algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the inter-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum [28].
- *Bag of Visual Words*. Another important algorithm for image feature detection is BoVW [18]. The general idea of BoVW is to represent an image as a set of features. These features consist of key-points and descriptors. Key points are the points in an image; so, no matter how much the image is rotated, shrunk, or expanded, its key points will always be the same. The descriptor is the description of the key point. We used the key points and descriptors to construct vocabularies and represented each image as a frequency histogram of features that are in the image. From the frequency histogram, we can find other similar images or predict the category of the image. To calculate key points and descriptors, we used ORB [29].
- *Programming instruments*. We used Python programming language, as it is one of most popular programming languages for image classification. Image processing

and computer vision algorithms were implemented by using open source library OpenCV [30]. While Scikit-Learn was our main library during machine learning training [31], Keras and Tensorflow were used during deep learning training [15].

5 Data Collection and Dataset Formation

5.1 Data Collection

- *Ethics and regulations*. We followed all ethics and regulations during data collection. All parents were properly informed of test regulations and their consents were obtained. Students were guided by their teachers during process. It was an extra-curricular activity for students with minimum level of stress.
- *Participants*. Students from private educational institution were selected. They were aged from 4 to 11 years, from Nursery 1–2 and Primary 1–2–3–4 grades. More than 1000 pictures were collected, and some meaningless sketches were eliminated, resulting with almost 1000 pictures remaining.
- *Process*. Students were given a plain sheet and told to sketch a man. No further assistance was given as the sketch required the authentic work of the child. About 10–15 min were given to the children to finish the job. After the sketching process was completed, all pictures were collected and passed for further steps.
- *Manual classification*. A consensus of three trained persons (one school counselor, one PhD student and one Physiology Professor) followed Goodenough's 51-point scoring criteria to calculate the score of every picture (Goodenough, 1926). Once scoring was done, we classified them according to mental ages, from 4 to 11 years, resulting a total of 8 classes. Below, in Table 2 we present total numbers of images for every class.

Table 2. Number of images in each class for All_8 dataset.

Class	4	5	6	7	8	9	10	11
No of images	142	131	152	160	125	124	79	38

5.2 Dataset Formation

To make our work diverse and to find best possible option, we created three versions of the dataset.

a) *All_8*: this is primer dataset, all sketches divided to 8 classes, according to the ages, from 4 to 11 years. Pictures were cropped and resized to 120 × 240 pixels.
b) *Double*: in this dataset, we merged the adjacent age groups; 4 with 5, 6 with 7, 8 with 9 and 10 with 11. This resulted in a total of 4 classes.
c) *Reference*: this is small dataset, where we joined classes 10 and 11. This dataset was used during the CKP experiment to propose a novel method.

The formation of different datasets is not a desirable condition and our expectation is to get the best possible result on All_8 dataset. However, during the experiment we discovered that the classification of All_8 is a very difficult task, which forced us to look for different methods of classification.

As a rule of thumb, we used 75% of the data for training and the remaining 25% for testing. We used training and testing dataset for other classification methods. In CKP, you do not need such partition of data. In CKP, we need only Reference dataset for formation of vocabulary and rest of data was used for testing.

6 Methodology

6.1 Counting Key Points (CKP) Algorithm

Firstly, we followed the process described in Fig. 6. During each experiment we have chosen different values for X and Y, to see how good the prediction accuracy would be. So, the set of pairs we selected for (X, Y) = {(200, 50), (800, 250), (1000, 400), (2000, 800)}.

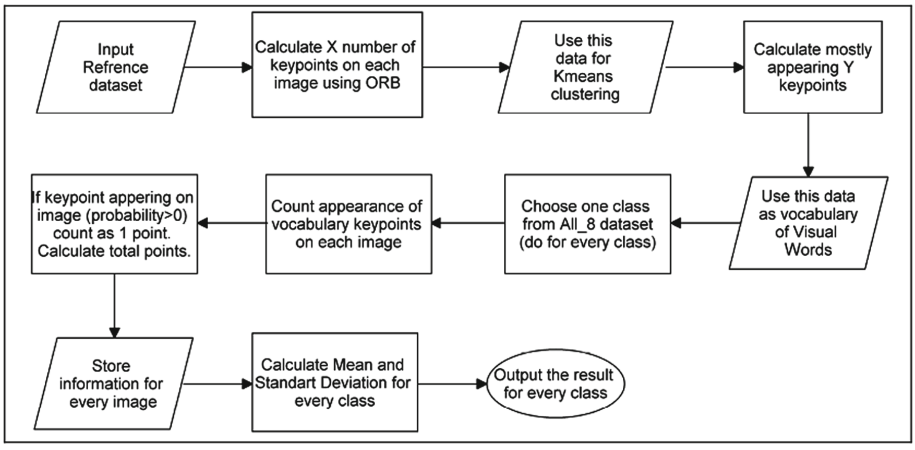

Fig. 6. CKP algorithm process scheme, first stage.

For instance, assume that (X, Y) = (1000, 400). We proceed with followings steps:

1. Step1. First, we used the Reference dataset to calculate 1000 key points and their descriptors from each picture, using ORB. Figure 7 displays some of selected images and their respective key points (1000) with red circles.
2. Step2. Next, we used K-means to find the 400 most likely descriptors to appear in every picture. We formed a vocabulary of visual words from these 400 descriptors.
3. Step3. We used matching method to determine if these vocabulary elements are appearing in other classes of All_8, and how frequently they appeared. If a descriptor was appearing on picture (probability of appearing is bigger than zero) we counted it as 1 point.

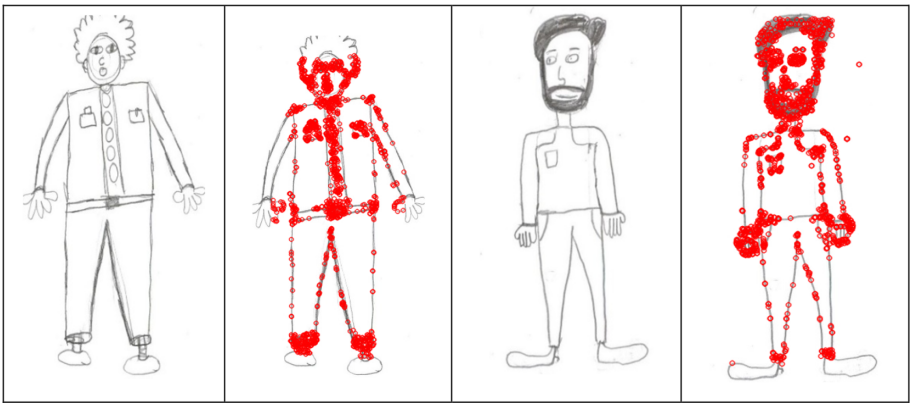

Fig. 7. Sample pictures and their key-points for BoVW.

4. Step 4. We calculated the total points for each picture.
5. Step 5. Lastly, we produce summary of the mean and standard deviation for every class (rounded to integer).

Table 3 is summary for the operation presented on Fig. 6 (after Step 5) for every set of (X, Y). The experiment is done on All_8 dataset. Regardless of the selected set, we can observe that there is difference between means of the classes and it is gradually increasing with every class.

Table 3. Mean and standard deviation for each class and set for All_8.

		Set 1	Set 2	Set 3	Set 4
Class	Stats.	(200; 50)	(800; 250)	(1000; 400)	(2000; 800)
4	*Mean*	35	95	113	134
	Std.	5	27	37	53
5	*Mean*	35	109	135	161
	Std.	4	26	37	54
6	*Mean*	37	114	141	174
	Std.	7	24	36	56
7	*Mean*	37	121	152	187
	Std.	4	22	34	52
8	*Mean*	38	122	158	195
	Std.	3	20	33	57
9	*Mean*	38	132	170	217
	Std.	3	19	30	53
10	*Mean*	38	139	183	250
	Std.	3	20	32	64
11	*Mean*	40	145	192	268
	Std.	3	15	30	62

Hypothetically, we can assume that we can classify the sketches with respect to their means. But, biggest challenge in this case is the standard deviation, it is very high. Next, we applied same operation on Double dataset. Table 4 is summary of the experiment on Double dataset.

Table 4. Mean and standard deviation for each class and set, for Double.

		Set 1	Set 2	Set 3	Set 4
Class	Stats.	(200; 50)	(800; 250)	(1000; 400)	(2000; 800)
45	*Mean*	35	101	118	147
	Std.	4	27	36	55
67	*Mean*	37	118	141	181
	Std.	4	23	36	54
89	*Mean*	38	127	156	206
	Std.	4	20	35	56
1011	*Mean*	39	141	180	256
	Std.	3	18	33	64

If we compare results on Tables 3 and 4, we can observe that average standard deviation for every set of (X, Y) is not changing, but still very high. Definitely, this will cause a serious challenge during prediction operation. At this point, we strongly encourage some further researches in order to reduce the value of standard deviation to improve classification accuracy. But, the margins between means are wider in Table 4, which would directly affect the classification accuracy in positive manner.

Next, we aimed to use information from Tables 3 and 4 to develop an image classification methodology. Figure 8 is process chart to develop this methodology.

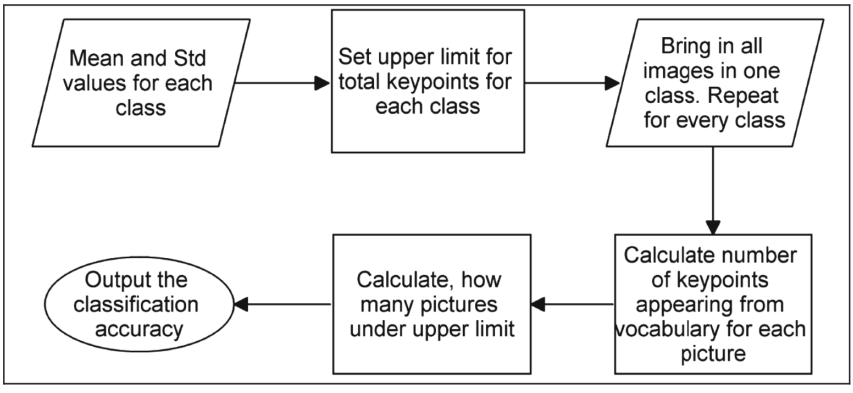

Fig. 8. methodology for development of prediction of image class

As standard deviation is too high for every class, we were careful with selection of upper limit for every class. The limits should be reasonable with respect to mean and deviation of two neighbor classes. Table 5 is the summary and limits we have chosen for every class in All_8 and Double dataset.

The rest of operation in Fig. 8 can be completed and described with algorithm presented in Fig. 9. This operation should be done for every class and accuracy should be recorded.

Table 5. Upper limit for each class in both All_8 and Double dataset.

	All_8								Double			
Set No.	4	5	6	7	8	9	10	11	45	67	89	1011
Set 1 (200; 50)	34	35	36	37	38	39	40	40+	35	37	38	38+
Set 2 (800; 250)	102	112	118	123	128	136	142	142+	109	123	132	132+
Set 3 (1000; 400)	127	138	145	155	163	177	188	188+	130	148	168	168+
Set 4 (2000; 800)	145	165	181	191	206	229	264	264+	164	194	231	231+

Algorithm Counting Keypoints

```
A ← all images in class B
C ← upper limit for class B
t ← total number of images in class B
0 ← sum
for every image in A do
    D ← Count total key point appearance form vocabulary
    if D ≤ C then
        sum ← sum + 1
accuracy ← sum/t
return accuracy
```

Fig. 9. Counting key-points pseudocode

Finally, we can present prediction accuracies in Tables 6 and 7. We present prediction accuracies for each class and total prediction accuracy, which is weighted accuracy of the set. We can observe that Set 4 (2000; 800) is giving best result on both All_8 and Double dataset. As it was expected, prediction accuracy on Double dataset is better than All_8, and we managed to reach highest 65% of prediction accuracy. We could keep on this version and produce Set 5 (4000; 1200) or Set 6 (6000; 2000), but this didn't increase the accuracy, instead it led to overfitting the model. This may be applicable on very large datasets, but for our case Set 4 seems to be most optimal.

Table 6. Prediction accuracy for All_8 dataset.

Class	No of images	Set 1	Set 2	Set 3	Set 4
4	142	*0.42*	*0.64*	*0.63*	*0.62*
5	131	*0.49*	*0.48*	*0.52*	*0.56*
6	152	*0.46*	*0.5*	*0.52*	*0.55*
7	160	*0.44*	*0.45*	*0.52*	*0.55*
8	125	*0.45*	*0.6*	*0.58*	*0.59*
9	124	*0.44*	*0.5*	*0.51*	*0.58*
10	79	*0.54*	*0.42*	*0.47*	*0.56*
11	38	*0.1*	*0.65*	*0.63*	*0.5*
Total accuracy	**951**	***0.443***	***0.522***	***0.543***	***0.57***

Table 7. Prediction accuracy for Double dataset

Class	No of images	Set 1	Set 2	Set 3	Set 4
45	273	*0.42*	*0.57*	*0.6*	*0.65*
67	312	*0.45*	*0.5*	*0.56*	*0.63*
89	249	*0.4*	*0.55*	*0.6*	*0.67*
1011	117	*0.75*	*0.75*	*0.7*	*0.66*
Total accuracy	**951**	***0.465***	***0.564***	***0.6***	***0.65***

6.2 Comparison of CKP Algorithm with Other Classification Methods

We developed one-to-one models of two milestone research papers for classification of hand drawn sketches; Eitz et al [2] and Yu et al [4]. In brief, Eitz et al firstly calculated Histogram of Oriented Gradients for each sketch than feed this data to SVM, while Yu et al developed a Convolutional Neural Network model to classify hand drawn sketches. Table 8 is summary of classification accuracies for aforementioned two milestone models and CKP algorithm. We can clearly observe that other two methods struggled to understand the dataset, as it needs special approach, while CKP overperformed both.

Table 8. Prediction accuracies three different methods

Method	All_8	Double
SVM + HOG	*22%*	*48%*
CNN	*32%*	*52%*
CKP	*57%*	*65%*

7 Discussion of the Results

In this paper, we aimed to conduct some classification experiments on DAPT images and we presented the methodology for CKP algorithm. Throughout the experimentation, we discovered that this kind of images should use some specific way of approach, rather than common classification methods. We observed that the All_8 dataset is difficult to classify with high accuracy, while joining 2 classes like in Double, can minimize this burden. However, our primary goal should be to develop a method to classify All_8 dataset.

We introduced a novel method, CKP algorithm, which seems to be simple but achieved good results with this type of dataset. As we have tested a unique dataset, which was not tested before for classification by computational device, we believe that our resulted prediction accuracy, 65%, is worthy finding and can be improved with further studies. We also presented comparative results on Table 8, proving that CKP is superior with comparison to existing state-of-art sketch classification methods.

The main shortcoming of the algorithm is, as we stated several times before, the standard deviation of the classes is very high. Prediction accuracy might improve if future studies can lower the standard deviation.

References

1. Ha, D., Eck, D.: A neural representation of sketch drawings (2017). arXiv preprint arXiv:1704.03477
2. Eitz, M., Hays, J., Alexa, M.: How do humans sketch objects? ACM Trans. Graph. **31**(4), 1–10 (2012)
3. Schneider, R.G., Tuytelaars, T.: Sketch classification and classification-driven analysis using fisher vectors. ACM Trans. Graph. (TOG) **33**, 1–9 (2014)
4. Yu, Q., Yang, Y., Liu, F., Song, Y.-Z., Xiang, T., Hospedales, T.M.: Sketch-a-Net: a deep neural network that beats humans. Int. J. Comput. Vis. **122**, 411–425 (2017). https://doi.org/10.1007/s11263-016-0932-3
5. Rakhmanov, O.: Testing strength of the state-of-art image classification methods for hand drawn sketches. In: 2019 15th International Conference on Electronics, Computer and Computation (ICECCO), pp. 1–3. IEEE (2019)
6. Jolley, R.P.: Children and Pictures: Drawing and Understanding. Wiley (2009)
7. Goodenough, F.L.: Measurement of intelligence by drawings (1926)
8. Rey, A.: L'examen psychologique dans les cas d'encephalopathie traumatique. Arch. Psychol. **28**, 112 (1942)
9. Wolf-Klein, G.P., Silverstone, F.A., Levy, A.P., Brod, M.S., Breuer, J.: Screening for Alzheimer's disease by clock drawing. J. Am. Geriatr. Soc. **37**, 730–734 (1989)
10. Fastenau, P.S., Denburg, N.L., Hufford, B.J.: Adult norms for the Rey-Osterrieth Complex Figure Test and for supplemental recognition and matching trials from the Extended Complex Figure Test. Clin. Neuropsychologist **13**, 30–47 (1999)
11. Naglieri, J.A., MacNeish, T.J., Bardos, A.: Draw a person: screening procedure for emotional disturbance. dap: sped. Pro-Ed (1991)
12. Sun, Z., Wang, C., Zhang, L., Zhang, L.: Query-adaptive shape topic mining for hand-drawn sketch recognition. In: Proceedings of the 20th ACM International Conference on Multimedia, pp. 519–528 (2012)

13. Rakhmanov, O., Dane, S.: Correlation between vanderbilt ADHD diagnostic scale and the draw-a-man test in school children. J. Res. Med. Dent. Sci. **7**, 77–81 (2019)
14. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. SSS. Springer, New York (2009). https://doi.org/10.1007/978-0-387-84858-7
15. Gulli, A., Pal, S.: Deep Learning with Keras. Packt Publishing Ltd (2017)
16. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**, 91–110 (2004)
17. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) 2005 Jun 20, vol. 1, pp. 886–893. IEEE (2005)
18. Yang, J., Jiang, Y.-G., Hauptmann, A.G., Ngo, C.-W.: Evaluating bag-of-visual-words representations in scene classification. In: Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval, pp. 197–206. ACM, New York, NY, USA (2007). https://doi.org/10.1145/1290082.1290111
19. Baxes, G.A.: Digital Image Processing: Principles and Applications. Wiley, New York (1994)
20. Li, Y., Hospedales, T.M., Song, Y.-Z., Gong, S.: Free-hand sketch recognition by multi-kernel feature learning. Comput. Vis. Image Underst. **137**, 1–11 (2015)
21. Yu, Q., Yang, Y., Song, Y.-Z., Xiang, T., Hospedales, T.: Sketch-a-Net that Beats Humans (2015). arXiv:1501.07873 [cs]
22. Ballester, P., Araujo, R.M.: On the performance of GoogLeNet and AlexNet applied to sketches. In: Thirtieth AAAI Conference on Artificial Intelligence (2016)
23. Xu, P., Joshi, C.K., Bresson, X.: Multi-graph transformer for free-hand sketch recognition (2019). arXiv preprint arXiv:1912.11258
24. Bensalah, A., Riba, P., Fornés, A., Lladós, J.: Shoot less and sketch more: an efficient sketch classification via joining graph neural networks and few-shot learning. In: 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), pp. 80–85. IEEE (2019)
25. Yama, M.F.: The usefulness of human figure drawings as an index of overall adjustment. J. Pers. Assess. **54**, 78–86 (1990)
26. Naglieri, J.A., McNeish, T.J., Achilles, N.: Draw a person test. In: Tools of the Trade: A Therapist's Guide to Art Therapy Assessments, p. 124 (2004)
27. Rakhmanov, O., Agwu, N.N., Adeshina, S.A.: Experimentation on hand drawn sketches by children to classify Draw-a-Person test images in psychology. In: The Thirty-Third International Flairs Conference (2020)
28. Tan, P.-N.: Introduction to Data Mining. Pearson Education, India (2018)
29. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.R.: ORB: An efficient alternative to SIFT or SURF. In: ICCV, p. 2. Citeseer (2011)
30. Gary, B., Adrian, K.: Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly Media, Inc. (2008)
31. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)