Tilburg centre for Creative Computing
Tilburg University
http://www.uvt.nl/ticc

P.O. Box 90153
5000 LE Tilburg, The Netherlands
Email: ticc@uvt.nl

April 26, 2009

TiCC TR 2009–002

A New Benchmark Dataset for Handwritten Character Recognition

Laurens van der Maaten
TiCC, Tilburg University

Abstract

The report presents a new dataset of more than $40,000$ handwritten characters. The creation of the new dataset is motivated by the ceiling effect that hampers experiments on popular handwritten digits datasets, such as the MNIST dataset and the USPS dataset. Next to a character labeling, the dataset also contains labels for the 250 writers that wrote the handwritten character, which gives the dataset the additional potential to be used in forensic applications. The report discusses that data gathering process, as well as the preprocessing and normalization of the data. In addition, the report presents the results of initial classification and visualization experiments on the new dataset, in an attempt to provide a base line for the performance of learning techniques on the dataset.

A New Benchmark Dataset for Handwritten Character Recognition

Laurens van der Maaten
TiCC, Tilburg University

1 Introduction

The availability of large collections of digitized, preprocessed, and labeled (image) data is of high importance to researchers in machine learning, since such datasets allow researchers to appropriately assess the performance of their learning techniques. Even though the UCI repository provides a large number of such datasets, many of these datasets do not contain sufficient data instances to facilitate good comparisons between techniques for machine learning. Two notable exceptions are the MNIST and USPS handwritten digits datasets [3, 9]. Over the last decade, a large number of studies in machine learning have used these two datasets to evaluate the performance of their approaches. Examples of approaches that have been evaluated on the MNIST include shape features [1], various flavours of Support Vector Machines [2, 9, 8], a variety of feed-forward, deep-belief, and convolutional networks [5, 9, 11, 17], and dimensionality reduction techniques [6, 13, 20].

The popularity of the MNIST (and USPS) dataset is the result of two main characteristics of the dataset. First, the number of datapoints in the dataset is, in contrast to many other publicly available datasets, sufficiently large, but the entire dataset can still easily be stored in the RAM of a modern computer. Second, models that are trained on the MNIST dataset can often be evaluated intuitively, because humans are typically very good at the identification handwritten digits. For instance, human observation of the receptive fields learned by Restricted Boltzmann Machines may provide insight in the quality of the procedures that were used to train the model [4].

An important problem of experiments with the MNIST dataset is that the experiments typically suffer from a severe ceiling effect. For instance, the generalization error of a simple 1-nearest neighbor classifier on the official test set is approximately 3% [9]. If the data is deskewed and blurred before training the classifier, the generalization error of 1-nearest classifiers trained on the MNIST dataset even drops to 1.22%. Even if it is not allowed to add characters to the training data that have small elastic or affine distortions, the most sophisticated classifiers have trouble beating this score. To date, the best models achieve generalization errors on the test set between 0.5% and 1.0%. However, it is likely that these models suffer from overfitting on the test data due to the optimization of their parameters with respect to the generalization error on the test set. Given this severe ceiling effect, the value of evaluations of machine learning models on the MNIST dataset is debatable.

In this report, we present a new handwritten character dataset that defines a more difficult classification problem than the MNIST dataset, because it includes both handwritten digits and characters. The new dataset consists of $40,121$ images of handwritten characters tha comprise 35 classes: 25 upper-case character classes and 10 digit classes (the character 'X' is not included as a class in the dataset). This report describes the way in which the dataset was collected, and presents the results of initial classification and visualization experiments on the new dataset. In our experiments, we found most classifiers had a generalization of between 15% and 20%, as a result of which it is not likely that future experiments on the new dataset are hampered by ceiling effects.

The outline of the remainder of the report is as follows. In Section 2, we discuss the way in which the dataset was collected, including the collection and digitization of the handwritings and the character segmentation, normalization, and labelling. Section 3 presents the results of some initial classification and visualization results on the new dataset. Some concluding remarks are presented in Section 4.

NADAT ZE IN NEW YORK, TOKYO, QUEBEC, ZÜRICH EN OSLO WAREN GEWEEST, VLOGEN ZE UIT DE USA TERUG MET VLUCHT KL658 OM 12 UUR. ZE KWAMEN AAN IN DUBLIN OM 7 UUR EN IN AMSTERDAM OM 9.40 UUR 'S AVONDS. DE FIAT VAN BOB EN DE VW VAN DAVID STONDEN IN R3 VAN HET PARKEERTERREIN. HIERVOOR MOESTEN ZE HONDERD GULDEN (*f* 100,-) BETALEN.	AFTER THEY HAD BEEN IN NEW YORK, TOKYO, QUEBEC, ZÜRICH, AND OSLO, THEY FLEW BACK FROM THE USA ON FLIGHT KL658 AT 12 O'CLOCK. THEY ARRIVED IN DUBLIN AT 7 O'CLOCK AND IN AMSTERDAM AT 9.40 O'CLOCK IN THE EVENING. THE FIAT OF BOB AND THE VW OF DAVID WERE IN R3 ON THE PARKING LOT. FOR THE PARKING, THEY HAD TO PAY HUNDRED GUILDERS (*f* 100,-).

Table 1: Dutch text that the subjects were instructed to copy (in uppercase script), and the translation of the text in English.

2 Dataset

In this section, we present the new handwritten character dataset. We discuss the gathering of the handwritings and the preprocessing of the gathered data separately in Section 2.1 and 2.2.

2.1 Handwriting collection and digitization

The original handwriting images were collected by Schomaker and Vuurpijl [16] in 2000, and was published as a dataset for forensic writer identification, called the *Firemaker* dataset. The Firemaker dataset consists of Dutch text handwritten by 251 students. Each student wrote four different pages: (1) a page with a specified text in natural handwriting, (2) a page with a specified text in uppercase handwriting, (3) a page with a specified text in 'forged' handwriting, and (4) a page with a free text in natural handwriting. For the collection of the characters dataset, we only used the second pages.

The handwriting pages contain a Dutch text of 612 alphanumeric characters in uppercase script. All 251 different students were instructed to write exactly the same text. The text that the students were instructed to copy was specifically designed to contain a variety of alphanumeric characters, making it well suited for the construction of a handwritten character dataset. The Dutch text, as well as its English translation, is printed in Table 1. On the response sheets that were used by the students, lineation guidelines were provided using a dropout color. A dropout color is visible by the human eye, but fully reects the light emitted by the scanner lamp such that is has the same appearance as the white background. In the collection of the handwritings, the recording conditions were standardized: all students used the same kind of paper and the same pen. The handwriting pages were scanned in grayscale on a flatbed scanner at 300 dpi with 8 bits per pixel.

2.2 Character collection and labeling

In the collection of the handwritten characters from the digitized pages with uppercase handwriting, we employed a semi-automatic approach to segmentation of characters from the handwriting[1]. Specifically, we invert and binarize the handwriting pages, and perform a labeling algorithm on the resulting image in order to label the binary objects in the handwriting image. Subsequently, we remove binary objects that are too small to form a complete character (objects with less than 8 pixels), and we present the binary objects to a human labeler. The human labeler has the option of rejecting the segmentation, for instance, when a part of the character is missing in the segmented binary object. If the human labeler rejects a segmented binary object, the binary object is simply ignored. If the human labeler accepts the segmented character, the human labeler manually labels the character. As the text that the writers were instructed to write is known by the human labeler, the labeling of the characters is unambiguous. For instance, the human labeler is thus capable of distinguishing between 'O' and '0' characters.

[1] Automatic segmentation of handwritten characters is an unsolved problem because of Sayre's paradox: a character cannot be segmented before it has been classified and it cannot be classified before it has been segmented [14, 22].

After the segmentation, we normalized the (labeled) handwritten character images by resizing them in such a way that they fit into a box of 50×50 pixel box, and the aspect ratio of the characters is retained. Subsequently, we performed two different kinds of normalization on the handwritten characters: (1) center-of-box normalization, and (2) center-of-mass normalization. The two types of normalization give rise to two different datasets, which we call dataset A and B. The two types of normalization are described below.

In the center-of-box normalization, a handwritten character is placed in a square box (of fixed size) in such a way that the number of border pixels on top and below the character are equal, and that the number of border pixels left and right from the character are equal. On all sides of the character images, empty borders with a width of 3 pixels are added resulting in the final 56×56 pixel character images. The resulting normalized images are stored in dataset A.

In the center-of-mass normalization, a character is placed in a box in such a way that the center of mass of the object is located in the center of the box. The box in which the characters are centered has size 90×90 pixels. The resulting images are stored in dataset B.

The character collection process described above has led to a dataset that contains $40,121$ handwritten

Figure 1: Randomly selected handwritten characters from dataset A.

characters, which are stored using the two types of normalization (in dataset A and B). Moreover, the dataset contains a set of with the $40,121$ original unedited character images, a list of character labels, and a list of writer labels. The writer labels were automatically stored, which was possible because each writer wrote on a different page (and thus in a different image), and may be used in forensic application. In Figure 2.2, we present 100 randomly selected handwritten characters from dataset A (center-of-box normalized). Even though the text that was written by the human writers was designed in such a way as to contain all characters, the class distribution of the resulting dataset is skewed, because, for instance, vowels occur more often in natural language than consonants. The class distribution of the new dataset is shown in Figure 2.2. Note that the character 'X' is missing in the dataset, as it was not part of the text that the writers were instructed to write.

3 Experiments

In the previous section, we presented the digitization and the preprocessing of the handwritten characters dataset. In this section, we present the results of classification and visualization experiments we performed on the new dataset (on the characters of dataset B). The main aim of these experiments is to provide a basis for comparisons with future state-of-the-art machine learning techniques. In the classification experiments, we trained and tested k-nearest neighbor classifiers on the two datasets, as well as

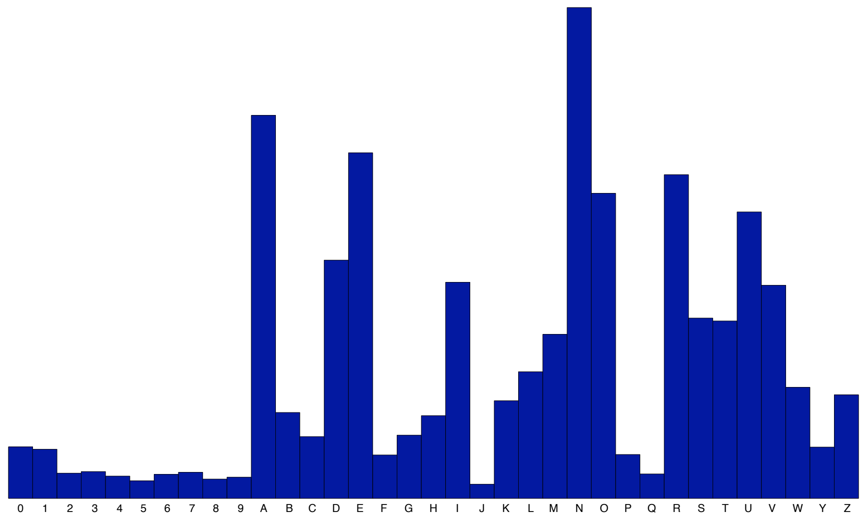


Figure 2: The class distribution in the handwritten characters dataset.

the recently proposed linear kernel classifiers (LKC) that use an approximation to the kernel trick based on random features [10]. The employed linear kernel classifiers were trained using a one-versus-all scheme, and used a Gaussian kernel with a bandwidth σ that was optimized using cross-validation. In the visualization experiments, we constructed visualizations using PCA, Isomap [18], Locally Linear Embedding [12], and t-SNE [21].
We present the results of the classification experiments in Section 3.1, and the results of our visualization experiments in Section 3.2.

3.1 Classification experiments

In Table 2 and 3, we present the results of experiments in which we measured the generalization error of a variety of classifiers on the new handwritten characters dataset (both dataset A and B). The reported generalization errors were measured using 10-fold cross-validation, and are reported with their corresponding standard deviation. We did not deskew and/or blur the characters in the dataset, but it is likely that doing so would improve the obtained generalization errors.
The results in Table 2 and 3 reveal that, in contrast to experiments on the MNIST dataset, experiments on the new handwritten characters dataset are not hampered by a ceiling effect. The lowest generalization error we measured (on dataset B) was 17.23%. The reader should note that this error is slightly misleading, for instance, because distinguishing between the characters 'O' and '0' is generally only possible when the context of the character is available, which is not the case in our dataset. In order to provide more insight into which errors our classifiers make, we present the confusion matrix for one of the experiments in Table 4, specifically, for our experiment with the 1-nearest neighbor classifier. The confusion matrix in Table 4 reveals that many of the errors are the result of the classifier confusing 'O' and '0'. If this type of confusions is ignored (i.e., if an instance with class 'O' may also be labeled as '0' and vice versa), the generalization error of the 1-nearest neighbor classifier is still 13.34%.

It is well known that the performance of classifiers on handwritten character recognition tasks can be improved by adding slightly distorted copies of the original data to the training data [9]. These distortions may include small translations, rotations, rescalings, or other affine transformations.

3.2 Visualization experiments

In addition to the classification experiments, we performed experiments in which we use dimensionality reduction techniques to reduce the $56 \times 56 = 3136$ dimensions of the image data in dataset A to two

Classifier	*Generalization error*
1-NN	$21.68\% \pm 0.61\%$
3-NN	$20.79\% \pm 0.44\%$
5-NN	$20.74\% \pm 0.64\%$
LKC	$32.99\% \pm 0.74\%$

Table 2: Generalization errors and the corresponding standard deviations of classifiers that were trained on handwritten characters dataset A (measured using 10-fold cross validation).

Classifier	*Generalization error*
1-NN	$17.79\% \pm 0.53\%$
3-NN	$17.23\% \pm 0.82\%$
5-NN	$17.28\% \pm 0.66\%$
LKC	$30.20\% \pm 0.55\%$

Table 3: Generalization errors and the corresponding standard deviations of classifiers that were trained on handwritten characters dataset B (measured using 10-fold cross validation).

dimensions. We present the results of the dimension reduction in a scatter plot, in which the colors correspond to the character labels of the images. In Figure 3, we present the results of the experiments with Principal Components Analysis [7] (which is identical to classical scaling [19]), Isomap [18], Locally Linear Embedding (LLE) [12], and t-Distributed Stochastic Neighbor Embedding (t-SNE) [20]. The results indicate that most dimensionality reduction techniques are not well capable of modeling the class structure of the data in the constructed two-dimensional maps.
Since t-SNE seems to have produced the best visualization, we also present a t-SNE visualization in which the original handwritten character images are plotted back in the visualization (in Figure 4). The resulting plot provides insight into (some of) the natural clusters in the dataset, and also reveals that it may not be easy for a machine learner to successfully separate the classes in the dataset (some of the natural clusters are overlapping in the visualization).

4 Concluding remarks

We introduced a new dataset of handwritten characters, which was created in order to address ceiling problems from which experiments on the popular MNIST and USPS datasets suffer (both in classification and visualization experiments). The new dataset also has writer labels, facilitating its use in forensic applications (for instance, in the construction of grapheme codebooks [15]). We presented some initial experiments in which we evaluated the performance of classifiers trained on the handwritten characters dataset, and in which we constructed visualizations of the new dataset. The results of the experiments indicate that the current dataset is more challenging than the MNIST and USPS datasets, as a result of which it facilitates better comparisons between state-of-the-art techniques for handwritten character recognition.

Acknowledgements

The author thanks Louis Vuurpijl and Lambert Schomaker for generously providing the Firemaker dataset. Laurens van der Maaten is supported by NWO/CATCH, project RICH, under grant 640.002.401.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	Y	Z
0	13	0	0	0	0	0	0.22	0	0	0	0	0	0.65	6.1	0	0	1.1	0	0.22	0	0	0	0	0	75	0	0	0	0	0	2.6	1.1	0	0	0.22
1	0	21	0	0	0	0	0	0.23	0	0	2.7	0	0	0	0	0	0	0	75	0	0	0.23	0	0	0	0	0	0	0	0.46	0	0	0	0.23	0
2	0	0	41	0	0	0	0	0.45	0.45	0	0.89	0	0.45	0.45	0	0	0	0	0.89	0	0	1.8	0	0.45	0	0	1.8	0.89	0	0	0	0	0	0	51
3	0	0	0.42	67	0	2.1	0	0.84	0	0.42	0	2.1	0	0.84	0	0	0	0	0.42	1.3	0.42	0	0	0	0	0	4.6	0	16	1.3	0	0	0	0.42	2.1
4	0	0.5	0	0	56	0	0	0	0	2	1	0	2	0	0	0.5	0.5	2.5	0.5	0	0	2.5	0	0.5	1.5	0	3	0	1.5	0	15	2	0.5	8	0
5	0.64	0	0	0.64	0	27	0	0	0	0	0	0	0.64	0	0.64	0.64	1.9	0	0	0	0	0	0	0	0.64	0	2.6	0	63	1.3	0	0	0	0	0
6	0.47	0	0	0.47	0	0	47	0	0	0	0	0	5.2	0.47	1.4	0	23	0	0.47	0	0	12	0	0	2.8	0	3.8	0	1.4	0	0.47	0.95	0	0	0
7	0	0.86	0	0	0	0	0	72	0	1.3	0	0	0	0	0	0.43	0	0	4.7	0	0	0	0	0	0	1.3	3.4	0	0	9.1	0	1.3	0	3.9	2.2
8	0.58	0	1.2	0	0	2.3	0	0	22	0.58	0.58	2.9	0	1.7	2.3	0	0.58	0.58	1.7	1.2	0	0	0.58	0.58	12	0	4.7	0	41	0	0	0	0	0	3.5
9	0	0.54	0	1.1	6.5	0	0	2.7	0.54	42	0	0	0	1.1	0	0	2.2	0.54	1.1	0.54	0	0	0.54	0	1.6	0	11	0	11	0	0.54	0.54	0	17	0
A	0	0.59	0	0	0.15	0	0	0.03	0	0.03	95	0	0.03	0.21	0	0.03	0	0.53	0.06	0	0	0.12	0.15	0.97	0.12	0.12	1.3	0.09	0.03	0	0	0	0	0.12	0.03
B	0.39	0.26	0.13	1.2	0	0.26	1.8	0	1.4	0.26	1.3	54	0.92	6.7	3	0	5.6	0.66	0.39	0.13	0	0.92	0.66	0.26	3.9	0	4.1	1	6	0.13	0.26	0.13	0	0.13	3.8
C	0	0	0	0	0	0	0.55	0	0	0	0	0	90	0	2.6	0	0.37	0	0.18	0	0	6	0	0	0.37	0	0.37	0	0	0	0	0	0	0	0
D	3	0	0	0	0	0	0.24	0	0	0	0.33	0.14	0.19	70	0	0.05	0.09	0.05	0.14	0.05	0	1	0	0.99	18	1.8	2	0.05	0.38	0	0.09	0.66	0	0	0
E	0	0	0	0	0	0	0.23	0	0	0	0	0.03	5.4	0	90	0.81	0.07	0	0.23	0	0.03	1.3	0	0	0.13	0	0.88	0.07	0.1	0.07	0	0	0	0	0.2
F	0	0.78	0	0	0.26	0	0	0.78	0	0	0.26	0	0	0	1.3	68	0	0	1.3	0	0.52	0.78	0	0	0	1.6	2.3	0	0	22	0	0	0	0	0
G	3	0	0	0	0.7	0	6.3	0	0	0.35	0.17	0	2.6	0	0.87	0	58	0	0	0.35	0	0.52	0	0	17	0	5.6	0	2.8	0	1.4	0.35	0	0.17	0
H	0	0	0	0	1.4	0	0	0.14	0	0.14	1.4	0	0	0	0	0	0	75	0	0	0.54	0.27	13	3.1	0	0.14	3.3	0.14	0	0	1.4	0.82	0.14	0	0
I	0.05	16	0.1	0.1	0	0	0	0.1	0	0	0	0	0.05	0	0.1	0.1	0	0	81	0.52	0	0.57	0	0	0	0	0.05	0	0.1	0.63	0	0	0	0	0.05
J	0	0.79	0	0	0	0	0	0.79	0	0	0	0	0	0	0	0	0	0	7.9	72	0	0	0	0	0	0	10	0	7.1	0	0.79	0	0	0.79	0
K	0	0.46	0	0	0.12	0	0.12	0	0	0	0.23	0	0.69	0	0.12	0.23	0	0.35	1.3	0.12	73	8.2	0.23	0.23	0	0.12	3.7	2	0	0	3.3	4.5	0.23	0.69	0.12
L	0	0.44	0	0	0	0	0	0	0	0	0	0	2.2	0	0.27	0	0	0	0.88	0	0	96	0	0	0	0	0.35	0	0	0	0	0.18	0	0	0
M	0	0.14	0	0	0.34	0	0	0.14	0	0	0.27	0	0	0.27	0	0	0	8.2	0.07	0.14	0	0.07	78	6.2	0	0.07	4.3	0	0	0.41	0.69	0.48	0.27	0.27	0
N	0.02	0.04	0	0	0	0	0.07	0	0	0	0.46	0	0	0.09	0	0.02	0	0.75	0.27	0	0	0.32	0.55	95	0.11	0.09	1.2	0.02	0	0.02	0.07	0.75	0.39	0.05	0
O	12	0	0	0	0	0	0.18	0	0	0	0.04	0	0.29	3.7	0.07	0	1.1	0	0.04	0	0	0	0	0.04	80	0	0.81	0	0.11	0	1	0.59	0	0	0
P	0.26	0	0	0	0	0	0	0.26	0	0	0	0	0	6.9	0.26	3.1	0	0	0	0	0	0.51	0.26	1.5	1.5	74	8.5	0.77	0	1.3	0	0.51	0	0	0
Q	4.6	0.46	0	0	0.93	0	0.46	0	0	0	0.93	0	0.46	0.93	0	0	4.2	0	0.93	0.93	0	0.46	0	1.4	30	0.93	41	0.93	0.46	0.46	7.4	1.9	0.46	0	0
R	0.04	0.14	0.14	0	0.04	0	0	0.07	0	0	0.8	0.04	1.8	0.45	1.1	0.14	0.07	0.24	0.21	0.04	2.5	0.52	0.07	0.9	0.17	1.7	1.3	86	0	0.04	0.07	0.45	0.04	0.07	0.38
S	0	0	0	0.12	0.12	1.6	0.25	0	0.06	0	0.06	0.06	0.06	0.19	0.06	0	0.25	0	0.5	1.5	0	0	0	0	0.06	0	2.2	0	93	0.12	0	0.06	0	0	0
T	0	0.38	0	0	0	0	0	0.7	0	0	0	0	0	0	0	0.32	0	0	1	0	0	0.06	0	0.06	0	0	0.7	0.06	0	97	0	0	0	0.06	0
U	0.2	0	0	0	0.59	0	0.08	0	0	0	0.04	0	0	0	0	0	0.04	0.31	0.04	0.16	0.28	0.04	0.04	0.16	1.1	0	1.1	0	0	0	90	5.8	0.2	0.04	0
V	0.05	0	0	0	0	0	0	0	0	0	0	0	0	0.05	0	0	0	0	0.16	0.21	0.05	0.21	0.05	0.16	0.32	0	0.84	0	0	0	5.6	91	0.26	0.79	0
W	0	0.2	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0.1	0.2	0	0.1	0	0	7.3	0.3	0	6.1	0	0	0	1.3	2.7	82	0	0
Y	0	0	0	0	4.6	0	0	0.44	0	3.3	0.44	0.22	0	0	0	0	0.22	0.44	1.5	0.44	0.22	0	0.44	0	0.44	0.22	7.9	0	0.44	0.66	0.44	4.4	0.22	73	0
Z	0	0.43	13	0	0	0	0	1.6	0	0	0.22	0	0.76	0	0.65	0.11	0	0	2.4	0.33	0	2	0	0	0	0.11	3.3	0.43	0	0.33	0	0	0	0.11	75

Table 4: Confusion matrix of our experiment with 1-nearest neighbor classifiers on character dataset B (measured using 10-fold cross validation).

References

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2001.

[2] D. DeCoste and B. Schölkopf. Training invariant support vector machines. *Machine Learning*, 46(1-3):161–190, 2002.

[3] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning*. Springer-Verlag, 2001.

[4] G.E. Hinton. To recognize shapes, first learn to generate images. In T. Drew P. Cisek and J. Kalaska, editors, *Computational Neuroscience: Theoretical Insights into Brain Function*. Elsevier, 2007.

[5] G.E. Hinton, S. Osindero, and Y. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554, 2006.

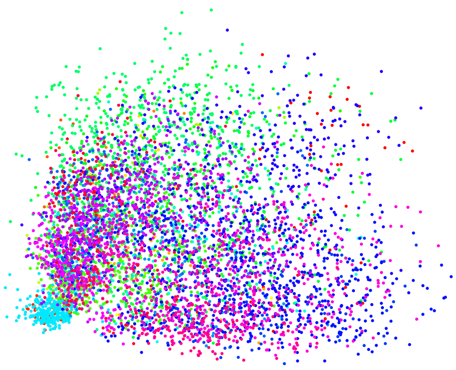
[6] G.E. Hinton and R.R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[7] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.

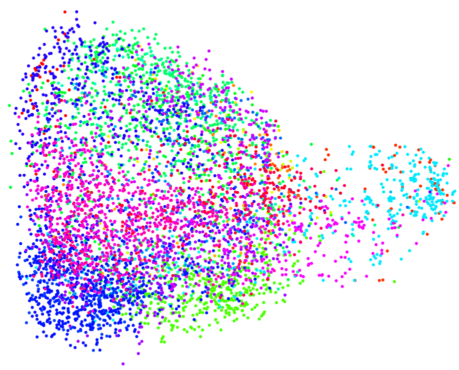
[8] F. Lauer, C.Y. Suen, and Gérard Bloch. A trainable feature extractor for handwritten digit recognition. *Pattern Recognition*, 40(6):1816–1824, 2007.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

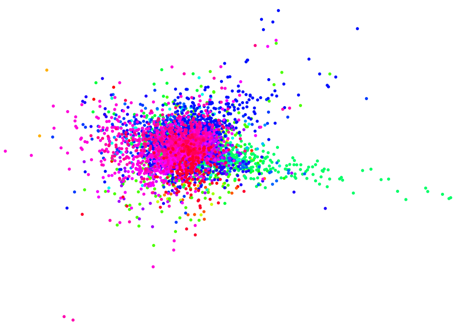
[10] A. Rahimi and B. Recht. Random features for large-scale kernel machines. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances of Neural Information Processing Systems*, volume 20, pages 1177–1184, Cambridge, MA, 2008. MIT Press.



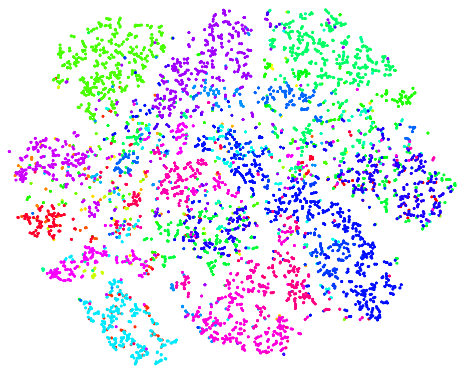
(a) Visualization constructed by PCA.



(b) Visualization constructed by Isomap.



(c) Visualization constructed by LLE.



(d) Visualization constructed by t-SNE.

Figure 3: Visualizations of the handwritten characters dataset.

[11] M.A. Ranzato, F.-J. Huang, Y.-L. Boureau, and Y. LeCun. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 1–8, 2007.

[12] S.T. Roweis and L.K. Saul. Nonlinear dimensionality reduction by Locally Linear Embedding. *Science*, 290(5500):2323–2326, 2000.

[13] R.R. Salakhutdinov and G.E. Hinton. Learning a non-linear embedding by preserving class neighbourhood structure. In *Proceedings of the* 11^{th} *International Conference on Artificial Intelligence and Statistics*, volume 2, pages 412–419, 2007.

[14] K. Sayre. Machine recognition of handwritten words: a project report. *Pattern Recognition*, 5(3):213–228, 1973.

[15] L.R.B. Schomaker, K. Franke, and M. Bulacu. Using codebooks of fragmented connected-component contours in forensic and historic writer identification. *Pattern Recognition Letters*, 28(6):719–727, 2007.

[16] L.R.B. Schomaker and L.G. Vuurpijl. Forensic writer identication: A benchmark data set and a comparison of two systems. Technical report, Nijmegen Institute for Cognition and Information, University of Nijmegen, The Netherlands, 2000.

[17] P.Y. Simard, D. Steinkraus, and J. Platt. Best practice for convolutional neural networks applied to visual document analysis. In *Proceedings of the International Conference on Document Analysis and Recogntion*, pages 958–962, 2003.

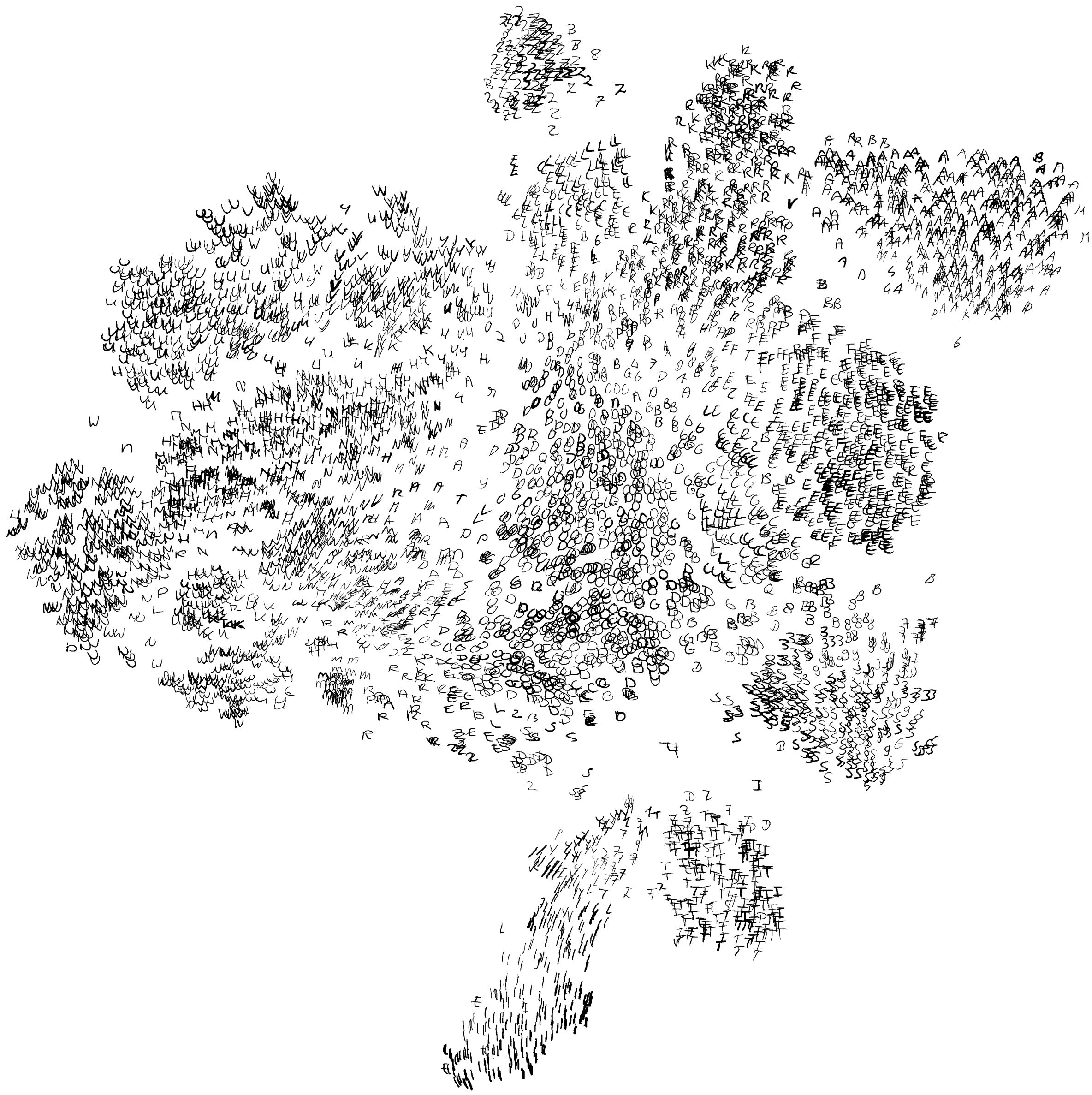


Figure 4: Magnified plot of the t-SNE visualization of the handwritten characters dataset.

[18] J.B. Tenenbaum, V. de Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[19] W.S. Torgerson. Multidimensional scaling I: Theory and method. *Psychometrika*, 17:401–419, 1952.

[20] L.J.P. van der Maaten and G.E. Hinton. t-Distributed Stochastic Neighbor Embedding. *Journal of Machine Learning Research*, 9 (in press), 2008.

[21] L.J.P. van der Maaten, E.O. Postma, and H.J. van den Herik. Dimensionality reduction: A comparative review, 2008.

[22] A. Vinciarelli. A survey on off-line cursive word recognition. *Pattern Recognition*, 35(7):1433–1446, 2002.