Applied Modelling in Economics,
Finance and Social Sciences,
2021

# Replication of the Performance of Image Classification Algorithms.

Kiril S. Kirilov
Sofia University “St. Kl.Ohridski”
kiril.st.kirilov@gmail.com

Online conference, 2021

# Image Classification / Face Recognition, some specifics:

- If colorful, the images are greyscaled and the pixels' values are normalized in the interval [0, 1], whereas a value of zero represents pure black and a value of one pure white color, respectively;
- For computational complexity reduction the images are usually copped (the $s \times t$ pixels' matrices are transformed as $p \times q$ whereas $s << p$ and $t << q$);
- The obtained $p \times q$ - pixels matrices are transformed as $p \times q$ - dimensional vectors;

Models, a performance replication of which we conduct:

SPP [1], DSNPE [2], SLPDP [3], DSPP [4];

# *Expectations:*

We expect that the performance of the algorithms in our experiments follows their chronological appearance in the literature. Also, in the original paper where the examined algorithms are introduced for the first time, they are compared with one or more previous models we consider, stating that the productivity of the new proposed algorithm exceeds the previous ones. Prior the run of the algorithms we expect the following effectiveness order:

SPP < DSNPE < SLPDP < DSPP

# *Databases:*

1. The public face images database ORL: ORL contains 400 grey scale images of 40 persons with a resolution of 112 x 92 pixels. Each person has 10 images that have variations in the lightning, facial expressions (smiling/not smiling, open/closed eyes) and/or other details (glasses/no glasses);

2. The AR face database: AR contains over 4000 color face images of 126 individuals (70 men and 56 women), including frontal views of faces with different facial expressions, illumination conditions and occlusions. The pictures of most persons were taken in two sessions (separated by two weeks). Each section contains 13 color images. In our experiments here, we use a subset of the AR face database provided and preprocessed by Martinez [5]. This subset contains 1400 face images corresponding to 100 person (50 men and 50 women), where each person has 14 different images with illumination change and expressions. The original resolution of these image faces is 165 x 120 pixels;

# I. ORL database - example:

Example of the images of person 1 from the ORL database,
dim = [112 x 92]

# *I. ORL database - experiments, prerequisites and experimental design:*

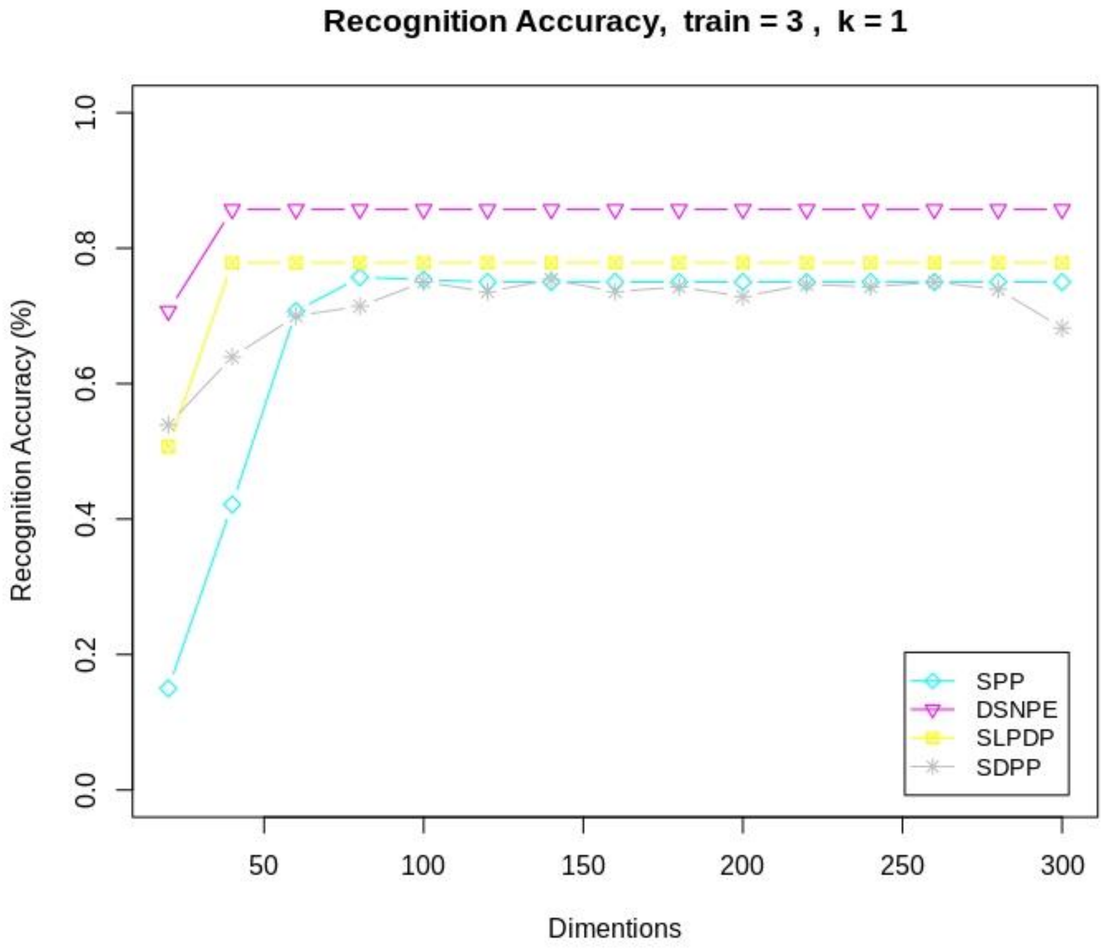
- The images are manually cropped to 32 x 32 pixels;
- Two scenarios with a random training sample of **3** and **5** images per person, respectively (30% and 50%);
- A PCA is performed on the training set with a 98% retention of the images’ energy;
- SPP, SPNPE, SLPDP and SDPP are performed for a subspace learning over the defined training subsets;
- The test samples are projected over the generated subspaces and evaluated with the application of 1-NN classifier;
- Parameters: the parameter $\varepsilon$ in the sparse representations has the value of 10 and for the parameter $\gamma$ in SPNPE we take the value of 1 (as in [2]) and the value of 0.1 in SLPDP (as in [3]), respectively;

# *I. ORL database – example of the cropped images:*

Example of three original and cropped images of one person from the ORL database, dim = [32 x 32]

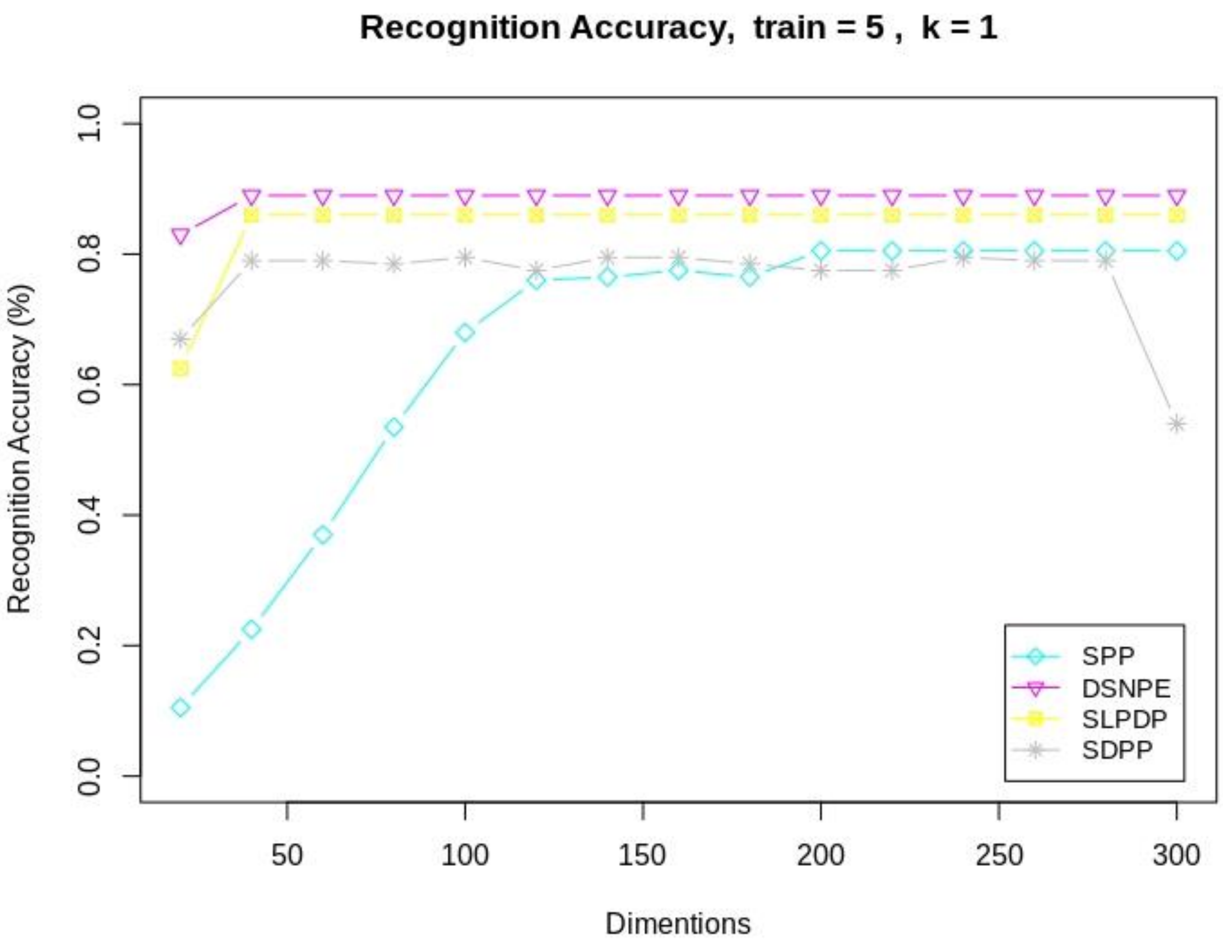
# I. ORL database – RA results:

Recognition accuracy (%) vs. number of projected vectors of the SPP, DSNPE, SLPDP and SDPP algorithms on the ORL database with a sample of **3 images** per subject for training (30% training set).



# *I. ORL database – RA results:*

Recognition accuracy (%) vs. number of projected vectors of the SPP, DSNPE, SLPDP and SDPP algorithms on the ORL database with a sample of **5 images** per subject for training (50% training set).



# *I. ORL database – projections:*

A sample of 3 images from the ORL database that will be projected on the best (DSNPE) and worst (SDPP) performing algorithms, trained over **3** images per subject.

# *I. ORL database – projections:*

The same 3 images projected on the *best* performing algorithm (DSNP), trained over **3** images per subject.

# *I. ORL database – projections:*

The same 3 images projected on the *worst* performing algorithm (SDPP), trained over **3** images per subject.

# *II. AR database - example:*

Example of the images of person 1 from the AR database,
dim = [165 x 120 x colour]

## *II. AR database - experiments, prerequisites and experimental design:*

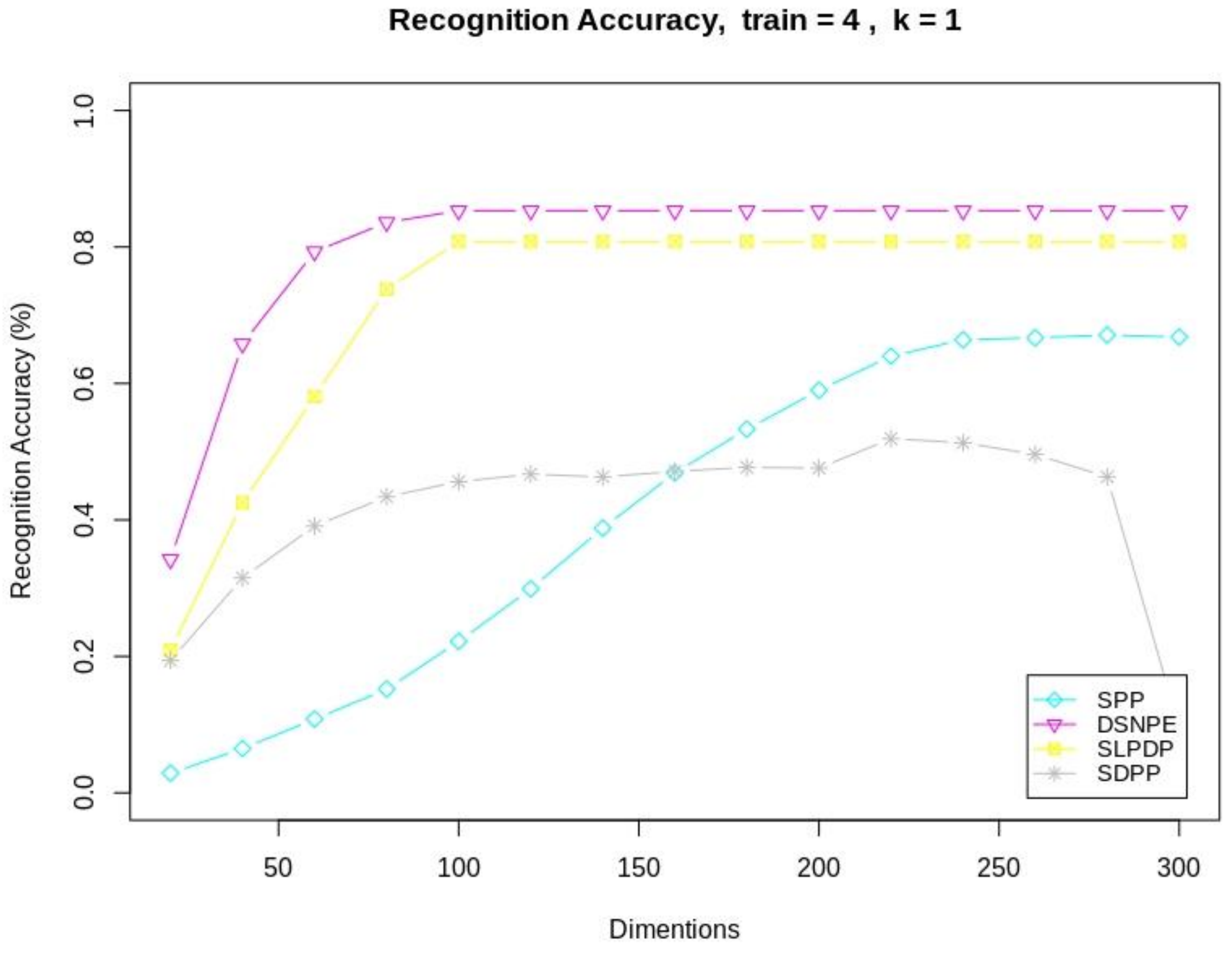
- The images are greyscaled and manually cropped to 32 x 32 pixels;
- Two scenarios, with a random training sample of **4** and **7** images per person, respectively (28.5% and 50%);
- A PCA is performed on the training set with a 98% retention of the images' energy;
- SPP, SPNPE, SLPDP and SDPP are performed for a subspace learning over the defined training subsets;
- The test samples are projected over the generated subspaces and evaluated with the application of 1-NN classifier;
- Parameters: For the sparse representations we use the MSR II, eq. (16) from [1]; For the parameter $\gamma$ in SPNPE we take the value of 1 (as in [2]) and the value of 0.1 in SLPDP (as in [3]), respectively;

# *II. AR database – example of the cropped images:*

Example of three
original (greyscaled)
and cropped images of
one person from the
AR database,
dim = [32 x 32].

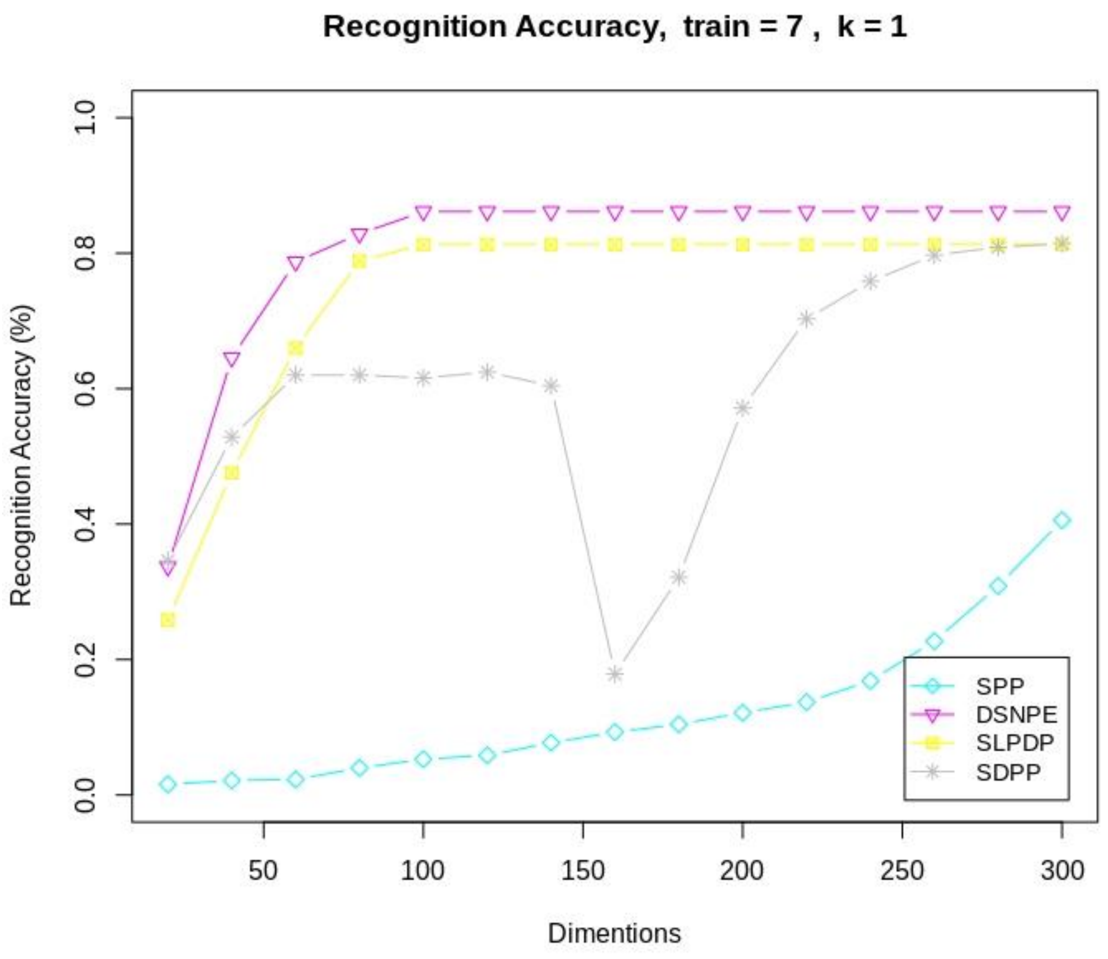
# *II. AR database – RA results:*

Recognition accuracy (%) vs. number of projected vectors of the SPP, DSNPE, SLPDP and SDPP algorithms on the AR database with a sample of **4 images** per subject for training (28.5% training set).



# *II. AR database – RA results:*

Recognition accuracy (%) vs. number of projected vectors of the SPP, DSNPE, SLPDP and SDPP algorithms on the AR database with a sample of **7 images** per subject for training (50% training set).



# *II. AR database – projections:*

A sample of 3 images from the AR database that will be projected on the best (DSNPE) and worst (SDPP) performing algorithms, trained over **4** images per subject.

# *I. ORL database – projections:*

The same 3 images
projected on the *best*
performing algorithm
(DSNPE), trained over
**4** images per subject.



# *I. ORL database – projections:*

The same 3 images projected on the *worst* performing algorithm (SDPP), trained over **4** images per subject

# Experimental results, comments:

- During the conduction of our experiments, we observe some interesting results in terms of the models' performance, that are contradicting our a priori expectations. In all four scenarios, we register a clear and stable advantage of the DSNPE algorithm over the remaining three models. In three, out of the four, cases the SDPP algorithm performs worst. Also, in the same three cases, the recognition accuracy order is the following DSPP < SPP < SLPDP < DSNPE, whereas in the fourth case it is: SPP < DSPP < SLPDP < DSNPE;
- The RA of the best algorithm doesn't increase a lot with the increase of the training set (from 30% to 50% for the ORL and from 28.5% to 50% for the AR database, respectively); This leads to the conclusion that the algorithms in the area are able to capture a significant amount of pattern and discriminant information also with small training sets;
- In all cases, the DSPP algorithm has a swinging RA curve, with a fall in the higher dimensions and a significant fall in the AR-7-train-case around the 150th-160th dimension. To observe the reasons for this could provide a useful information on the choice of an optimal projection vectors area;
- The performance of SPP falls dramatically during the significant increase of the AR training set from 4 to 7 images per subject. This confirms the statement that an increase of the training set may not necessarily ensure an addition of a discriminant information and opens a discussion on the choice of an optimal training set;
- A projections that is better recognizable to the naked eye doesn't necessarily mean that the respective algorithm performs better. This confirms the complex structures of images data.

# Conclusions:

In this work we replicate the performance of some famous algorithms for image classification and challenge the results in the original papers in which the models are introduced. The observations during the experiments we conduct contradict at a certain level the a priori expectations (that follow the algorithms chronology). This raises some questions on the:

1. Importance of the replication of the results of publications;
2. The requirement of a more descriptive and detailed disclosure of the experimental steps, designs, parameters etc., as their inconsistency and incomparableness may lead to desired, non-comparable and misrepresented results;

# Further work:

- To include other algorithms in the replication experiments and to apply them further on another public databases;
- To examine the reasons of the obtained general RA results;
- To examine the reasons of the ‘swinging’ SDPP performance and the fall of the performance of SPP on the AR database;

# References:


1. Lishan Qiao, Songcan Chen, Xiaoyang Tan, “Sparsity preserving projections with applications to face recognition”, Pattern Recognition, 43-1:331-341, 2010.
2. Jie Gui, Zhenan Sun, Wei Jia, Rongxiang Hu, Yingke Lei, Shuiwang Ji, “Discriminant sparse neighborhood preserving embedding for face recognition”, Pattern Recognition, 45-8:2884-2893, 2012.
3. Jianbo Zhang, Jinkuan Wang, Xi Cai, “Sparse locality preserving discriminative projections for face recognition”, Neurocomputing, 260:321-330, 2017.
4. J. Zhang, J. Wang and K. Zhan, “Sparse Discriminant Preserving Projections for Face Recognition”, at 3rd International Conference on Information Technology and Industrial Automation (ICITIA 2018), ISBN: 978-1-60595-607-7, 2018.
5. A.M. Martinez, R. Benavente, “The AR Face Database”, CVC Technical Report, 1998.

# Thank you!

The talk was supported by the project
80-10-93/25.03.2021 from
Sofia University “St. Kl.Ohridski”, Bulgaria