Contaminated Images Recovery by Implementing Non-negative Matrix Factorisation

Pengwei Yang[1], Chongyangzi Teng[1], and Jack George Mangos[1,2]

[1]*Faculty of Computer Science, The University of Sydney, Sydney NSW 2000, Australia*
[2]*Faculty of Medicine, University of New South Wales, Kensington NSW 2052, Australia*

Abstract

Non-negative matrix factorisation (NMF) has been extensively applied to the problem of corrupted image data. Standard NMF approach minimises Euclidean distance between data matrix and factorised approximation. The traditional NMF technique is sensitive to outliers since it utilises the squared error of each data point, despite the fact that this method has proven effective. In this study, we theoretically examine the robustness of the traditional NMF, HCNMF, and $L2, 1$-NMF algorithms and execute sets of experiments to demonstrate the robustness on ORL and Extended YaleB datasets. Our research indicates that each algorithm requires a different number of iterations to converge. Due to the computational cost of these approaches, our final models, such as the HCNMF and $L2, 1$-NMF model, fail to converge within the iteration parameters of this work. Nonetheless, the experimental results illustrate, to some extent, the robustness of the aforementioned techniques.

1 Introduction

The processing of, and automated learning from noisy image data is a pressing challenge in the field of machine learning. This is because there are an enormous number of sources of noisy image data which present bountiful opportunities for meaningful research. These include blurred handheld photography for image reconstruction and denoising, or a number of medical image streams with intrinsic noise, such as speckle noise in ultrasonographic organ imaging, the high signal-to-noise ratio in the scattered vessel fragments against white and grey matter in cerebral CT, or time constraints (as are often faced in hospitals) limiting the quality of MRI data [7] [9] [17] [18] [10]. As such, improving machine learning performance on noisy image datasets presents many opportunities for scientific and technological advancement.

Dimensionality reduction techniques can lead to "better predictions and more interpretable data" [16]. Non-negative matrix factorisation (NMF) is one group of methods used to address noisy image data. NMF methods involve the decomposition of multivariate data into smaller, or compressed, representative matrices [15]. In a data matrix consisting of multivariate n-dimensional data vectors, as below, the vectors are factorised into an n*r matrix W and an r*m matrix H, where r is typically smaller than both n and m. As such, each data vector v is approximated by a linear combination of the columns of W, weighted by the components of h: $v \simeq Wh$.

As the number of basis vectors in W is generally substantially smaller than the number in the original data matrix, good approximation of the original data is achieved only if underlying latent structures are identifiable [15]. However, in cases where such underlying latent patterns are present, and where the non-negativity requirement is satisfied, NMF demonstrates good performance compared with other methods, such as PCA or vector quantization, particularly where PCA and NMF are used together vs. PCA alone [15] [20].

This paper will test the robustness of the standard NMF, HCNMF and $L_{2,1}$-NMF algorithms to an image clustering task on two related datasets. Both datasets are greyscale headshots – the aim being to cluster these into groupings based on the individual being photographed. The ORL "Dataset of Faces" dataset contains 400 images of 40 subjects, with variation in lighting, facial expression, and facial details (including glasses); the background, facial positioning, and image size is uniform [14]. The Extended YaleB dataset is larger, containing 2414 images of 38 subjects; each subject is photographed in 9 different poses and under 64 illumination conditions. All images are manually aligned, and of uniform size [2].

2 Related Work

2.1 Standard NMF

NMF methods are used to capitalise on the non-negativity property of the datasets upon which they are used. As such, their use is limited to circumstances in which this property can be demonstrated, but they have previously been used in areas including text mining, hyper-spectral imaging, and genetic research [8]. The standard algorithm is based on the decomposition of the original data matrix into two lower-dimensional non-negative matrix factors, where the Euclidean distance between the product and the original data matrix is used to optimise the solution[8].

However, as the NMF algorithm only allows additive combinations, it is optimal when the dataset contains additive Gaussian noise, and therefore fails on highly corrupted or otherwise heterogeneously noisy datasets [8]. This, importantly for our paper, includes datasets of faces where the faces are obscured by glasses or scarves.

2.2 HCNMF

Many other variants of the NMF algorithm have been developed for different purposes. Hypersurface cost-based NMF, first proposed by Hamza and Brady, works by minimising the summation of hypersurface costs of errors. The benefit of the HCNMF method is that it produces an algorithm less sensitive to outliers than the $L2$-norm approach.

The aforementioned properties have meant the HCNMF function has been shown to outperform other methods (including NMF, PCA, cNMF, among others) in the factorisation of a spectral library [11], and in selecting differentially expressed genes and tumour classification [12]. However, one significant drawback to the HCNMF method is its computational demand. Because the algorithm utilises Armijo's rule-based line search, optimisation is time-consuming [8].

2.3 $L_{2,1}$-NMF

The standard NMF algorithm employs a Frobenius norm-based loss function, which makes it sensitive to noise, as described above. In order to address this sensitivity, the $L_{2,1}$ method[14] employs a loss function whereby the $L_{2,1}$ norm of the error matrix is minimised; this reduces the influence of outliers/noisy data points by inhibiting their significance in learning the subspace [8].

In reducing the sensitivity of the NMF algorithm to noisy data points, the $L_{2,1}$ method is more robust on heterogeneous datasets. As such, $L_{2,1}$-NMF can perform more efficiently than standard NMF when dealing with real-world data which contains noise. Furthermore, robust NMF such as $L_{2,1}$-NMF could be used for feature selection when applied to complex domains [6].

3 Methodology

3.1 Cost function and optimisation

3.1.1 Standard NMF

The objective function of the standard NMF algorithm is defined as:

$$min_{W \geq 0, H \geq 0}||V - WH||_F^2$$

where V is the set of input data vectors and W, H are the factor matrices. This objective function is typically solved using the multiplicative update rule (MUR) [15]. The iterative updating algorithm of standard NMF is shown below:

$$F_j k \Leftarrow F_j k \frac{(XDG^T)_{jk}}{(FGDG^T)_{jk}}$$

$$G_k i \Leftarrow G_k i \frac{(F^T XD)_{ki}}{(F^T FGD)_{ki}}$$

3.1.2 HCNMF

The objective function of HCNMF is defined as:

$$\sum_{ij}(\sqrt{1 + (V - WH)_{ij}^2} - 1)$$

where the cost function is defined as:

$$\delta(x) = \sqrt{1 + x^2} - 1$$

The cost function is differentiable and bounded, and that the cost function is quadratic when the argument is small and linear when the argument is large (as shown in Fig.1):
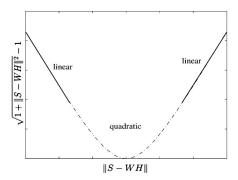

Figure 1: Loss Function of HCNMF [11]

The iterative updating algorithm of HCNMF is shown below:

$$W_{ik}^{(t+1)} = W_{ik}^{(t)} - \alpha_{ik}^{(t)} \frac{(WHH^T)_{ik}^{(t)} - (SH^T)_{ik}^{(t)}}{\sqrt{1 + ||S - WH||}}$$

$$H_{kj}^{(t+1)} = H_{kj}^{(t)} - \beta_{kj}^{(t)} \frac{(W^TWH)_{kj}^{(t)} - (W^TS)_{kj}^{(t)}}{\sqrt{1 + ||S - WH||}}$$

where alpha and beta are the step sizes opted at each iteration via Armijo rule for inexact line search [11]. Armijo rule is shown below [5].

$$f(x_k - \alpha_{mk}\nabla f(x_k)) - f(x_k) \leq -\frac{1}{2}\alpha_{mk}|\nabla f(x_k)|^2$$

3.1.3 $L_{2,1}$-NMF

The standard NMF utilises the Frobenius-norm-based loss function, which may mean that the the squared residual error might be large for any individual point. In order to handle that problem, $L_{2,1}$-NMF uses a robust formulation of the error function, which could be found as follows:

$$min_{W \geq 0, H \geq 0}||V - WH||_{2,1}$$

3

while the $L_{2,1}$-norm is defined as:

$$||E||_{2,1} = \sum_{j=1}^{n} ||E_j||_2$$

This method no longer employs the squared error, and hence why the $L_{2,1}$-NMF has better robustness to noise than standard NMF.

The iterative updating algorithm of standard $L_{2,1}$-NMF is shown below:

$$D_{ii} \Leftarrow \frac{1}{\sqrt{\sum_{j=1}^{p}(X - FG)_{ji}^2}} = \frac{1}{||x_- Fg_i||}$$

$$F_{jk} \Leftarrow F_{jk} \frac{(XDG^T)_{jk}}{(FGDG^T)_{jk}}$$

$$G_{kj} \Leftarrow G_{kj} \frac{(F^T XD)_{ki}}{(F^T FGD)_{ki}}$$

Researchers add a weighted matrix regulariser to the $L_{2,1}$-NMF, which aims to incorporate these weights to suppress outliers [13]. D is calculated as a diagonal matrix.

3.2 Noise

Real image data often contains noise. To compare how different NMF algorithms perform with different types of corruption, we simulated two typical types of noise: salt and pepper noise, and block-occlusion. We applied these types of noise to the two image datasets, ORL and Extended YaleB, to be used in our paper. Examples of these types of noise, including the varying degree to which they were applied, is shown in Fig.2.
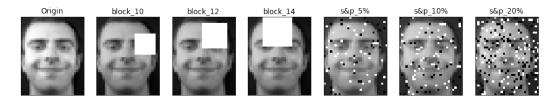

Figure 2: Demonstration of both block-occlusion noise (images 2-4) and salt-and-pepper noise (images 5-7) on an example image from the ORL dataset

3.2.1 Salt and Pepper (S&P) Noise

S&P noise is also known as impulse noise, and is defined as a sparsely occurring white and black pixel distribution superimposed on images [4]. It can be caused by several reasons, including defective camera sensors, or software or hardware failure in image capturing, or due to transmission/conversion error [3]. This type of noise will contaminate the image by a factor between zero and the maximum pixel value, and will do so in a random, pixel-by-pixel distribution. Typically the intensity value for pepper noise is close to 0, and close to 255 for salt noise:

$$\eta(x,y) = \{^{0,}_{255,} {}^{Pepper\,noise}_{Salt\,noise}\}$$

Of note, S&P noise can violate the Gaussian distribution noise assumption of standard NMF [4].

In this paper, S&P noise is simulated by randomly substituting a proportion p of the total image pixel count for any one image with either grayscale value 0 (black) or 255 (white). We have varied the percentage of corrupted pixels (p) to be 5%, 10% and 20% to demonstrate how the robustness of algorithms changes in response to higher amounts of noise. The proportion of salt to pepper noise is 0.5.

3.2.2 Block Occlusion

Image occlusion is also a common form of noise that can depreciate the performance of image recognition. This is a challenging type of noise as, in contrast to the dispersed noise of S&P, block occlusion obscures all information from one distinct region. This means that more traditional methods of overcoming S&P noise, such as standard or adaptive median filtration, are ineffective [1]. This type of noise, for a similar reason, is also challenging for NMF methods, which need to lean more heavily on dictionary information to account for the large amount of contiguous missing information.

In our paper, block occlusion is simulated by randomly placing a single b x b sized square block on each image. The block is given a uniform pixel greyscale value of 255. Three different block sizes were superimposed on distinct copies of the dataset, with b = 10, 12 and 14, to demonstrate how the robustness of tested algorithms changed with more information loss. The minimum block size and maximum block size imply 9% and 18% outlier for ORL images and 5% and 10% for Extended YaleB images.

3.3 Evaluation Metrics

We utilise three metrics for scoring algorithmic performance in each setting of noise.

3.3.1 RRE

The first is the relative reconstruction error (RRE). This metric measures the normalised distance between the clean dataset (Vhat) and the estimated dataset (WH) as an indication of the overall accuracy of the prediction, as below:

$$RRE = \frac{||\hat{V} - WH||_F}{||\hat{V}||_F}$$

3.3.2 Accuracy

The second is the average accuracy; this metric relates to the success of our algorithms in clustering the reconstructed images into their respective subject identities. This is calculated as the proportion of correctly clustered images compared to the overall sample size n, as below:

$$Acc(Y, Y_{pred}) = \frac{1}{n} \sum_{i=1}^{n} 1\{Y_{pred}(i) == Y(i)\}$$

3.3.3 NMI

Finally, normalised mutual information is used to evaluate clustering quality from the perspective of overall information gain, defined as the normalisation on the mutual information between cluster papers and the pre-existing input label. The normalisation used is the average of the entropy of the cluster paper and that of the input labelling[13], as below:

$$NMI(Y, Y_{pred}) = \frac{2 * I(Y, Y_{pred})}{H(Y) + H(Y_{pred})}$$

4 Experiment

4.1 Experimental Setup

The experiments were performed on the ORL and Extended YaleB datasets. To reduce the computational complexity, we have resized all ORL images to 30x37 pixels and all ORL images to 42x48 pixels.

Each of the ORL and Extended YaleB datasets was exposed to the same set of noise - clean, block 10, 12, and 14, and S&P 5, 10, and 20%. Three algorithms were then assessed for their robustness

- NMF, HCNMF, and the $L_{2,1}$-NMF. Each algorithm was tested on all proportions of noise for both datasets.

For rigorous performance evaluation, we have repeated each experiment 5 times by randomly sampling 90% of data from each whole dataset. The average and standard deviation of the metrics are reported.

4.2 Experimental Results

Table 1 and Figure 3 to 6 shows the average performance score of standard NMF, HCNMF and L2,1 NMF for both the ORL dataset and the extended YaleB dataset. Figure7 is a visual representation of the robustness of the models. The reconstructed images are produced by multiplication of matrices W and H that are learned based on the images with noise.

Table 1: Relative reconstruction error%, Accuracy% and NMI% with standard deviation

(a) ORL dataset

Noise Type	RRE			ACC			NMI		
	NMF	HCNMF	$L_{2,1}$	NMF	HCNMF	$L_{2,1}$	NMF	HCNMF	$L_{2,1}$
Clean	12.59±0.001	**12.38±0.018**	90.97±0.	71.56±0.026	**76.94±0.026**	73.89±0.018	84.22±0.012	**87.36±0.013**	85.18±0.007
block 10	**32.42±0.001**	32.43±0.001	91.16±0.	29.44±0.014	26.5±0.025	**30.17±0.021**	**46.88±0.008**	43.42±0.023	46.79±0.024
block 12	**38.49±0.001**	38.6±0.002	91.23±0.	22.72±0.008	**23.83±0.013**	22.94±0.004	41.45±0.01	**41.59±0.004**	40.32±0.007
block 14	**43.55±0.002**	43.59±0.002	91.24±0.	22.67±0.015	**24.06±0.016**	21.67±0.006	40.58±0.015	**41.23±0.015**	40.83±0.007
S&P 5%	**15.63±0.**	22.36±0.206	91.07±0.	69.83±0.023	54.78±0.073	**70.89±0.016**	82.26±0.014	65.35±0.045	**82.35±0.011**
S&P 10%	**18.8±0.001**	19.92±0.001	91.15±0.	65.17±0.022	**69.83±0.015**	66.28±0.023	78.56±0.007	**81.83±0.013**	78.83±0.015
S&P 20%	**24.22±0.001**	26.31±0.001	91.32±0.	52.44±0.015	**54.06±0.02**	52.28±0.019	68.17±0.006	**71.27±0.015**	68.36±0.018

(b) Extended YaleB dataset

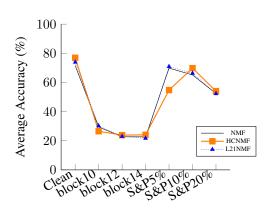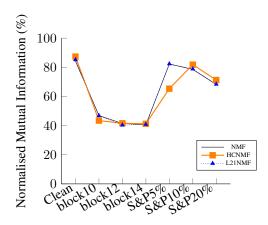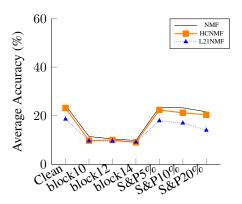| Noise Type	RRE			ACC			NMI		
	NMF	HCNMF	$L_{2,1}$	NMF	HCNMF	$L_{2,1}$	NMF	HCNMF	$L_{2,1}$
Clean	18.55(±0.001)	78.37±0.	81.59±0.002	24.29±0.011	9.24±0.003	18.56±0.006	32.58±0.016	9.6±0.006	25.59±0.009
block10	45.14±0.001	79.25±0.	86.59±0.	11.42±0.008	9.17±0.002	9.61±0.002	13.43±0.017	8.83±0.008	10.66±0.008
block12	53.72±0.001	79.53±0.	87.26±0.	10.43±0.009	9.±0.002	9.48±0.003	11.88±0.012	9.18±0.002	9.53±0.005
block14	62.59±0.002	79.8±0.	87.9±0.	9.81±0.003	9.16±0.003	9.11±0.002	11.08±0.007	8.74±0.004	9.17±0.004
S&P5%	19.86±0.001	78.88±0.001	85.14±0.	23.51±0.01	9.36±0.004	17.88±0.006	31.55±0.008	9.6±0.003	24.21±0.016
S&P10%	21.86±0.	79.21±0.001	86.29±0.001	23.24±0.008	9.27±0.001	16.94±0.009	31.39±0.013	9.47±0.006	22.93±0.014
S&P20%	26.99±0.001	79.74±0.	87.57±0.	21.61±0.012	9.01±0.001	14.±0.005	29.5±0.012	9.28±0.004	18.22±0.007

Figure 3: ORL - Accuracy

Figure 4: ORL - NMI

4.2.1 ORL dataset

All three models have similar performance. The accuracy and NMI on clean images are around 74% and 86% respectively. When we added salt&pepper noise, all three models still have satisfactory performance at 5% S&P corruption. As the corruption percentage increases, a decrease in the performance for all three models can be observed. For S&P 10% and 20%, HCNMF is slightly more robust. HCNMF has the worst performance at 5% corruption; this may be due to volatility, as the
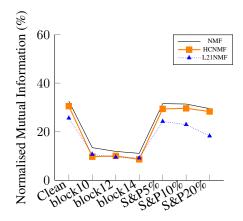
Figure 5: Extended YaleB - Accuracy
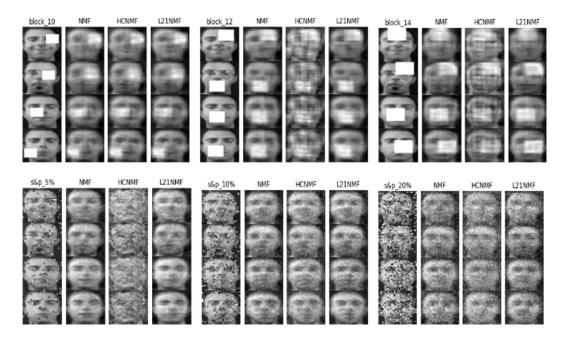

Figure 6: Extended YaleB - NMI

Figure 7: Reconstructed image - From left to right: Polluted image; Results by NMF; Results by HCNMF; Results by $L_{2,1}$-NMF

experiment is only repeated 5 times and outlier results could significantly skew the average score. This is evidenced by the higher standard deviation observed in table 1.

All three models have, unsurprisingly (given the challenging nature of the noise), a significant deterioration in performance with block occlusion. For a 10*10 block, which implies a 9% proportion of outliers for ORL images, the performance is significantly worse compared with a roughly equivalent salt and pepper noise of 10%.

The standard NMF converged before reaching the maximum iteration of 5000, however both HCNMF and $L_{2,1}$-NMF fail to converge. The results reported for HCNMF and $L_{2,1}$-NMF are based on their results at the max iteration steps.

4.2.2 Extended YaleB dataset

This dataset contains 2414 images and requires a significantly longer time to run an experiment. The optimisation of HCNMF and $L_{2,1}$-NMF is more difficult and time-consuming compared to standard

Table 2: Iteration impact demonstration

Step	Update diff of W	Update diff of H	Training time (mins)	RRE %	Accuracy %	NMI %
1000	0.0366	0.0705	0.87	91.3	53.0	68.4
5000	0.0362	0.0736	4.16	91.3	48.5	65.8
10000	0.0359	0.0748	8.17	91.3	49.0	66.1
20000	0.0358	0.0752	16.35	91.3	56.5	71.0

NMF. Due to the limited computational resources available, our team chose to cap the iteration steps at 3000 for HCNMF and 100 for $L_{2,1}$-NMF. At the max iteration both models have not converged to error bound, which has significantly impacted each respective model's performance.

When the iteration steps/running time is restricted, from the results table and chart, we can see the standard NMF outperformed both HCNMF and $L_{2,1}$-NMF.

Similar to what we observed from ORL data results, as image corruption increases the performance of the three NMF methods decreases. All three models perform better with respect to salt and pepper noise compared to block occlusion.

4.3 Reflection for future improvement

For future work, we would like to make a few improvements in the experiment design.

Firstly, increasing the max iteration cap to allow HCNMF and $L_{2,1}$ method to converge would potentially demonstrate their better performance on the two datasets. However, we were unable to facilitate this amount of testing due to limitations in time available and computational resources. As such, the reported results for these two methods are unconverged results. This has significantly impacted their performance and we are unable to fairly compare the robustness of the models relatively to each other.

To demonstrate that increasing the number of iterations would improve the performance when it gets closer to convergence point (global minimum point of the cost function), our team tested only the L2,1-NMF model using only 20% Salt and Pepper noise and a higher iteration step threshold. At the baseline number of iterations, the model does not converge with H and W updates differences above the threshold of 1e5 and continuing to decrease, however, as can be seen in Table 2, if the number of iteration steps is increased, the model performance demonstrates an improvement accordingly. Further experiments would require significantly longer run time due to the 5-fold cross-validation we've implemented, to trade off performance and run-time, we have set a lower threshold.

We would also like to test on a wider range of noise corruption. In this experiment, our team only tested three values for each type of noise. We have not compared the performance on extreme outliers, for example where more than 50% of the pixels are corrupted. In addition, due to limited points tested, we are not able to observe the rate of the performance deterioration and at which points the method starts to deteriorate significantly.

Lastly, to further improve the rigorousness of the performance evaluation, we would ideally repeat each experiment more than five times. However, increasing the number of repeat experiments will further increase the computational time.

5 Conclusion and Future work

In this paper, we implemented three types of NMF algorithms, namely standard NMF, HCNMF and $L_{2,1}$-NMF. We theoretically analysed the robustness of HCNMF and $L_{2,1}$-NMF. Considering the computational complexity of HCNMF and $L_{2,1}$-NMF methods, we chose to limit the number of update iterations, which meant that the models did not converge to the error bound at the cut-off point. As a result, our experiments have limited ability to demonstrate the robustness of both HCNMF and $L_{2,1}$-NMF. In the future, we could do further experiments to discover the range of convergence when establishing various models with ORL and YaleB dataset.

[19]

References

[1] Determining optimal window size for texture feature extraction methods. In: Sánchez, J.S., Pla, F. (eds.) Proceedings of the IX Spanish Symposium on Pattern Recognition and Image Analysis. Pattern recognition and image analysis, Publicacions de la Universitat Jaume, meeting Name: National Symposium on Pattern Recognition and Image Analysis

[2] The yale face dataset, `http://vision.ucsd.edu/content/yale-face-database`

[3] Alajlan, N., Kamel, M., Jernigan, E.: Detail preserving impulsive noise removal **19**(10), 993–1003. https://doi.org/10.1016/j.image.2004.08.003, `https://linkinghub.elsevier.com/retrieve/pii/S0923596504000761`

[4] Alqadi, Z.: Salt and pepper noise: Effects and removal **2**

[5] Armijo, L.: Minimization of functions having lipschitz continuous first partial derivatives. Pacific J Math **16**(1), 1–3 (1966)

[6] Díaz, A., Steele, D.: Analysis of the robustness of nmf algorithms. arXiv preprint arXiv:2106.02213 (2021)

[7] Fan, C.M., Namazi, N.M.: Image Motion Estimation from Blurred and Noisy Image Sequences. Proceedings of the 1998 International Conference on Image Processing, IEEE Computer Society, meeting Name: International Conference on Image Processing

[8] Guan, N., Liu, T., Zhang, Y., Tao, D., Davis, L.S.: Truncated cauchy non-negative matrix factorization **41**(1), 246–259. https://doi.org/10.1109/TPAMI.2017.2777841, `https://ieeexplore.ieee.org/document/8123883/`

[9] Gupta, N., Swamy, M., Plotkin, E.: Despeckling of medical ultrasound images using data and rate adaptive lossy compression **24**(6), 743–754. https://doi.org/10.1109/TMI.2005.847401, `http://ieeexplore.ieee.org/document/1435536/`

[10] Haldar, J.P., Jacob, M., Ebel, A., Zhu, X., Schuff, N., Hernando, D., Sutton, B.P., Liang, Z.P.: Regularised inversion of noisy, incomplete MR spectroscopic imaging data with anatomical prior. John Wiley, OCLC: 946274487

[11] Hamza, A., Brady, D.: Reconstruction of reflectance spectra using robust nonnegative matrix factorization **54**(9), 3637–3642. https://doi.org/10.1109/TSP.2006.879282, `http://ieeexplore.ieee.org/document/1677928/`

[12] Jiao, C.N., Gao, Y.L., Yu, N., Liu, J.X., Qi, L.Y.: Hyper-graph regularized constrained NMF for selecting differentially expressed genes and tumor classification **24**(10), 3002–3011. https://doi.org/10.1109/JBHI.2020.2975199, `https://ieeexplore.ieee.org/document/9004542/`

[13] Kong, D., Ding, C., Huang, H.: Robust nonnegative matrix factorization using l21-norm. In: Proceedings of the 20th ACM international conference on Information and knowledge management - CIKM '11. p. 673. ACM Press. https://doi.org/10.1145/2063576.2063676, `http://dl.acm.org/citation.cfm?doid=2063576.2063676`

[14] Laboratories, A.: The database of faces, `https://cam-orl.co.uk/facedatabase.html`

[15] Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorisation. Advances in neural information processing systems, MIT press

[16] Squires, S.E.: Non-negative matrix factorisation: algorithms and applications

[17] Wang, Y., Yan, G., Zhu, H., Buch, S., Wang, Y., Haacke, E.M., Hua, J., Zhong, Z.: VC-net: Deep volume-composition networks for segmentation and visualization of highly sparse and noisy image data **27**(2), 1301–1311. https://doi.org/10.1109/TVCG.2020.3030374, `https://ieeexplore.ieee.org/document/9222053/`

[18] Wu, D., Ren, H., Li, Q.: Learning-based CT perfusion image denoising with only noisy training data. Center for Advanced Medical Computing and Analysis

[19] Xia, X., Liu, T., Wang, N., Han, B., Gong, C., Niu, G., Sugiyama, M.: Are Anchor Points Really Indispensable in Label-Noise Learning? Curran Associates Inc. (2019)

[20] Zhao, L., Zhuang, G., Xu, X.: Facial expression recognition based on PCA and NMF. In: Proceedings of the 7th World Congress on Intelligent Control and Automation. `https://ieeexplore-ieee-org.wwwproxy1.library.unsw.edu.au/stamp/stamp.jsp?tp=&arnumber=4593968`