Proceedings of the ARW & OAGM Workshop 2019 DOI: 10.3217/978-3-85125-663-5-28

On the Use of Artificially Degraded Manuscripts for Quality Assessment of Readability Enhancement Methods*

Simon Brenner[1] and Robert Sablatnig[1]

Abstract— **This paper reviews an approach to assess the quality of methods for readability enhancement in multispectral images of degraded manuscripts. The idea of comparing processed images of artificially degraded manuscript pages to images that were taken before their degradation in order to evaluate the quality of digital restoration is fairly recent and little researched. We put the approach into a theoretical framework and conduct experiments on an existing dataset, thereby reproducing and extending the results described in the first publications on the approach.**

I. INTRODUCTION

Written heritage is a valuable resource for historians and linguists. However, the physical medium preserved may be in a condition that prohibits the direct access of the text: fading of the ink, darkening of the substrate or artificial removal due to substrate re-use (palimpsestation) are possible circumstances that render a text unreadable. Several imaging methods have been employed to recover such lost writings in the last fifteen years, with Multispectral and Hyperspectral imaging as well as X-Ray Fluorescence mapping being the most prominent base techniques [2], [7], [10], [11], [13], [20], [23], [24]. While the hardware systems are continually improved, post-processing methods for readability enhancement that were adapted for Multispectral Images of manuscripts over a decade ago [24] are still used by practitioners and prominently appear in recent literature [2], [10], [20]. Developments in the area are impeded by the absence of suitable metrics for automatically evaluating the quality of the results.

In literature describing methods for readability improvements of written heritage, evaluations based on expert ratings or the demonstration on selected examples is common [7], [13], [20], [23], [24]. This practice is unfavorable for the research field, as the evaluation of methods on large datasets is unfeasible if human assessment of results is required. Considering the fact that the development of computer vision methods typically involves multiple iterations of testing and improvement [31], the problem becomes even more apparent. A similar problem is faced by the practitioner, who is forced to manually try and visually evaluate a palette of methods in order to find the optimal result for a given investigated object.

*This work was supported by the Austrian Science Fund (FWF), grant no. P29892

[1]Simon Brenner and Robert Sablatnig are with the Computer Vision Lab, Institute of Visual Computing & Human-Centered Technology, TU Wien, 1040 Vienna, Austria `sbrenner@cvl.tuwien.ac.at`, `sab@cvl.tuwien.ac.at`

We propose that an ideal metric for the assessment of readability should have the following properties:

1) *Unsupervised.* The readability assessment does not require user input, such as selection of different pixel classes. Furthermore, a readability score can be calculated for unknown documents, i.e. documents where the contained text is not known a-priori.
2) *Culture agnostic.* The assessment is applicable to writings of any script and language equally.
3) *Consistent with expert ratings.* At the end of the day, domain experts still possess the highest authority for readability assessment, as it is them who then actually read the texts. Therefore, a ranking of a given set of enhancement results based on the calculated readability score should coincide with a ranking created by a domain expert.

Such a metric not only facilitates efficient testing and benchmarking, but also allows for optimization-based parameter tuning for postprocessing algorithms or the pre-selection of the best images from a large number of results from different algorithms.

A. Previous Approaches for Quantitative Evaluation

Several attempts for a quantitative assessment of text restoration quality are found in literature.

Arsene et al. [2] conducted a study on the effectiveness of a number of dimensionality reduction methods on a certain manuscript page. In addition to the obligatory score by expert rating, they employed the Davies-Bouldin Index and the Dunn Index, which are measures for cluster separability, as quality metrics. While all three metrics agreed on the best enhancement method, for the remaining positions of the ranking the computed scores diverged significantly from the human ratings, making their feasibility questionable. The authors acknowledge this and claim that the visual assessment by philologists is still the standard method of evaluating readability enhancement methods.

A natural assumption is that the quality of an image with regard to readability is strongly connected to its contrast. This is problematic however, as high contrast can be found in background noise and non-textual elements of a page (e.g. in the form of stains), especially when dealing with results of dimensionality reduction methods. Furthermore, the nominal contrast of an image can be increased by simple intensity transformations, thus rendering it impractical for the assessment of image quality. Faigenbaum et al. rely on the notion of *potential contrast* [26] to assess the readability of ostraca [8]. This measure rates the maximum contrast between foreground and background of a grayscale image that can be achieved by any intensity transformation. Although an intriguing idea, its implementation is problematic, as it relies on a binarization of the image by means of manually selected samples of foreground and background pixels, and the resulting score heavily depends on those samplings.

Another approach is to measure the quality of enhancement strategies by the performance of Optical Character Recognition (OCR) [14], [17]. In comparison to the preceding approaches, the evaluation via OCR performance has the advantage of directly related to the property of 'readability'. However, a ground truth is required and the results depend on the OCR algorithm employed and the data on which it was trained. Hollaus et al. [14], for example, evaluate their work on Glagolitic script and use a custom OCR system that has been trained for Glagolitic script only.

B. *Image Quality Assessment*

A closely related topic is the general Image Quality Assessment (IQA). Relevant approaches are categorized by the amount of information available to the estimator [5], [21].

Full-Reference (FR) methods have knowledge of a reference image that is assumed to be of optimal quality. The quality score is in essence a metric for the similarity between the reference image and a degraded version [1], [30]; a typical use case is the evaluation of lossy image compression, where an original image is naturally available.

No-Reference (NR) methods require no additional information aside from the input image that is to be evaluated. Successful NR IQA approaches, that are not limited to a certain type of distortion, typically employ machine learning one way or the other [4]. While early methods based on natural scene statistics, such as DIIVINE [22] or BRISQUE [21], are largely hand-crafted and just 'calibrated' on a training dataset, recent publications make heavy use of Convolutional Neural Networks (CNNs) [4], [5], [18], [15]. NR-IQA has been used to select optimal parameters for de-noising [21], [33] and artifact removal in image synthesis [3].

The problem of quantitatively evaluating readability enhancements can be considered a special case of IQA. For this application, however, a reference image is typically not available. It is thus natural that, using the taxonomy above, the assessment approaches outlined in Section I-A fall in the category of NR IQA (or Reduced-Reference IQA [30] in the case of evaluations based on OCR-performance). Although a NR approach would be preferable for the application, it is generally an ill-posed problem [18], even more so when focusing on the property of readability [10]. None of the approaches described above satisfies the requirements for an assessment metric we formulate. It is thinkable that CNN based approaches similar to those used for general NR IQA problems can be adapted and trained for readability assessment and used in a processing workflow for parameter optimization or pre-selection from a set of different results. For evaluation and benchmarking applications, however, CNNs are not a feasible option due to their dependence on a specific training process (which even introduces random components in the usual case of stochastic gradient descent optimization) [12] and the general opacity of their decision making [32].

C. *Artificial Degradation*

Giacometti et al. proposed a way to perform readability assessment in a FR setting [10]. They cut patches from an 18th century document written with iron gall ink on parchment and acquired Multispectral images before and after artificial degradation by various treatments. The resulting dataset [9] consists of 23 manuscript patches, of which 20 were subject to a different treatment each and three were left untreated as control images. Two of the patches were imaged from both sides, giving a total of 25 samples.

The dataset is then used to conduct a study on the performance of Multispectral imaging and postprocessing techniques in recovering information lost in the degradation process. The result images were compared with the untreated originals, which allows to view the approach as an instance of the FR IQA problem. The authors employ mutual information [29] as a similarity metric.

This work is of value and significance because to the best of our knowledge, it resulted in the first dataset systematically documenting the effects of degradation processes on the spectral response of written text and potentially enabling an objective evaluation of attempts to restore the original information. However, it has several restrictions for a broader application: First, the number of samples is small and, as all the samples are taken from the same manuscript, there is no variation in substrate and ink composition. Second, the important case of palimpsestation, i.e. the presence of a new layer of text on top of the degraded one, is omitted. Third, the accompanying paper [10] fails to conclusively show that comparison with the original image is a valid method to assess the quality of text restoration. Although plausible results are shown for selected examples, the generality of the results is not discussed; also it is not made clear which exact image is used for reference to obtain the specified mutual information scores. However, this is a prerequisite to legitimate further studies of this kind with a higher number of samples and greater variation.

In the following, we reproduce and extend the results described in the original paper in order to further investigate this third issue.

II. Continuative Experiments

The dataset described above contains multispectral images acquired with a monochromatic scientific camera as well as color images. In the following, we will only refer to the monochromatic images. For each sample, 21 spectral layers from 400nm to 950nm are available for the untreated and treated variants. The layers are intensity normalized [19] and inter-registered; however, the treated images are not registered to the untreated ones. Also a set of results from dimensionality reduction methods is provided for each sample; they are registered to the untreated variants, but far from pixel-accurately, prohibiting quantitative comparisons.

A. Preprocessing

For greater flexibility and accuracy we pre-processed the dataset prior to our experiments:

1) From the untreated image, a pan-chromatic image is created by averaging the layers in the visible range ($400nm < \lambda < 700nm$). For the sake of simplicity and uniformity, these panchromatic images will serve as a reference for registration and comparison, and will from here on be referred to as *reference*.
2) One layer of the treated sample is registered to the reference using a deformable registration framework for medical image processing [16], [25]. The 800nm layer was chosen for that purpose, as a visual assessment showed that it shares most of the textual information with the untreated images for the majority of degradation types. A deformable registration approach is necessary due to deformations of the parchment resulting from the treatments.
3) The remaining treated images are registered using the transformation found in the previous step.
4) Panchromatic images and registered treated images are cropped to 900x900 pixels.
5) To produce test images that can be compared with the reference, the cropped registered treated images are processed with five common (but arbitrarily chosen) dimensionality reduction methods: Principal Component Analysis (PCA), Independent Component Analysis (ICA), Factor Analysis (FA), Truncated Singular Value Decomposition (T-SVD) and K-Means Clustering (KM). From each method, five components were extracted, leading to a total of 25 processed variants for each sample, from here on referred to as *processed images*.

The three samples treated with heat, mold and sodium hypochlorite could not be registered satisfactorily due to their condition and were thus omitted, leaving 22 samples for investigation. The resulting modified version of the dataset is available online [6].

B. Comparison metrics

The images retrieved from dimensionality reduction methods visualize statistical dependencies rather than measured intensity values, such that contrast, mean brightness and *polarity* (in our case referring to dark text on light background versus light text on dark background) of these images typically deviate from the original photographs [10]. Therefore, any comparison metrics that rely on absolute intensity differences, such as the Mean Squared Error or Peak Signal To Noise Ratio, are unsuitable for this application. Instead, metrics that provide a measure of structural similarity and are insensitive to contrast and polarity are required.

Viewing the pixel positions as observations and the intensity values of the compared images as observed variables, statistical measures of dependence such as the Pearson Correlation Coefficient (PCC) and Mutual Information (MI) between the variables (i.e. images) are available as relevant comparison metrics. While MI, which Giacometti et al. employed in their work [10], can be used as-is, reversed polarities result in negative PCC vales such that the absolute value is used as a score.

Alternatively, established NR IQA metrics emphasizing structural similarity like the Structural Similarity Index (SSIM) [30] and Visual Information Fidelity (VIF) [27] are available. Although these metrics are not agnostic of contrast, its influence can be adjusted with a parameter for SSIM, while VIF actually rewards images with higher contrast than the reference. To make the methods invariant to polarity, we simply use $max(\varphi(I_{ref}, I_{test}), \varphi(I_{ref}, \neg I_{test}))$ as a comparison score, where φ denotes either SSIM of VIF between to images and $\neg$ is the image complement.

We consciously refrain from employing more advanced FR IQA metrics (e.g. based on learning) for these initial experiments as they would introduce unnecessary complexity.

C. Experiments

In order to reproduce previous results [10] and investigate the feasibility of comparison with an intact original as a measure for readability, we compare each processed image with the reference using MI as well as the adapted variants of PCC, SSIM and VIF described above. The use of additional similarity metrics allows to observe if the choice of metric significantly influences the results. The scores were then used to create rankings of the processed images for each sample, allowing to visually assess their plausibility.

In addition, the influence of contrast enhancement on the respective scores was experimentally evaluated: For each sample, the first five Principal Components (showing varying degrees of initial contrast) were subjected to Contrast Limited Adaptive Histogram Equalization (CLAHE) with varying clip limits, to monitor the influence on the different scores.

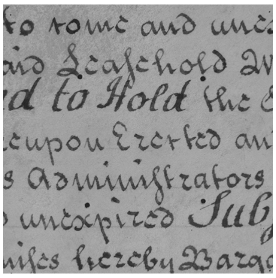
The full results of our experiments as well as relevant source code can be accessed online along with our preprocessed version of the dataset [6].

III. Discussion

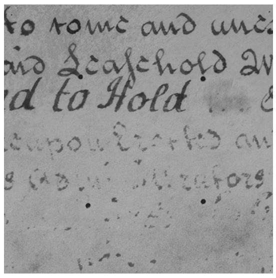
Visually assessing the processed image variants ranked by the employed comparison metrics generally confirms the assumption that similarity to a non-degraded reference image correlates well to the readability of text. The example shown in Figure 1 is representative for the remaining samples, where similar situations are observed.

The rankings derived from different similarity metrics are well correlated, with MI and PCC showing the strongest agreement. This is comprehensible when visually assessing rankings like in Figure 1c, and also manifests in the correlation matrix of the different metrics, which is shown in Table I.

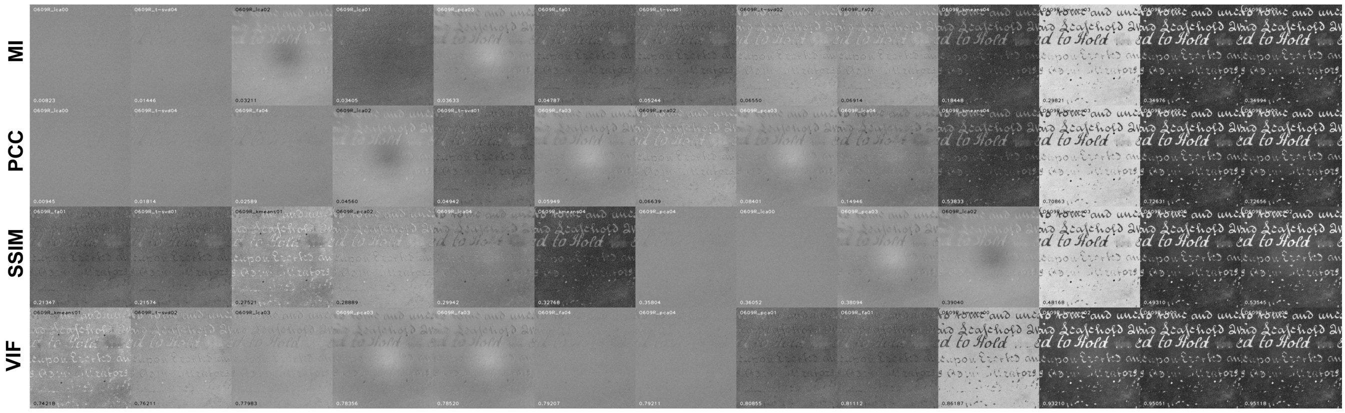
It might seem like the good scores of the highest ranked images are due to their high contrast; this general assumption, however, is readily disproved. Experiments with



(a) Untreated (panchromatic)



(b) Scraped (panchromatic)



(c) Ranked processed images

Fig. 1: An example of quality rankings derived from comparison with a reference image. (a) and (b) show panchromatic images of a sample of the dataset before and after artificial degradation via scraping. The rows of (c) correspond to the different metrics employed; the columns are ordered in ascending quality score. Due to space limitations, we only show every third column of the ranking.

	MI	PCC	SSIM	VIF
MI	1.0	0.9117	0.8189	0.7534
PCC	0.9117	1.0	0.8004	0.7211
SSIM	0.8189	0.8004	1.0	0.7395
VIF	0.7534	0.7211	0.7395	1.0

TABLE I: Correlation matrix of different employed similarity metrics, computed over all compared variants.

different levels of generic contrast enhancement showed that it has no positive effect on the scores. On the contrary, the SSIM and VIF scores decrease with increasing contrast. Figure 2 plots the mean deviations of similarity scores over the clip limit used for CLAHE contrast enhancements, along with the respective standard deviations. Note that the mean MI and PCC scores remain almost constant, whereby MI exhibits lower standard deviations. MI is thus the most stable of the tested metrics with respect to contrast alterations. The finding that generic contrast enhancements do not improve comparison scores is comprehensible, because the contrast of signal and noise is enhanced likewise. It also suggests that high comparison scores result from contrast that is also present in the original image (especially between text and foreground), which in turn supports the feasibility of image comparison as a quality metric for text restoration.

Although the results are visually convincing in general, individual examples for obviously erroneous ratings are

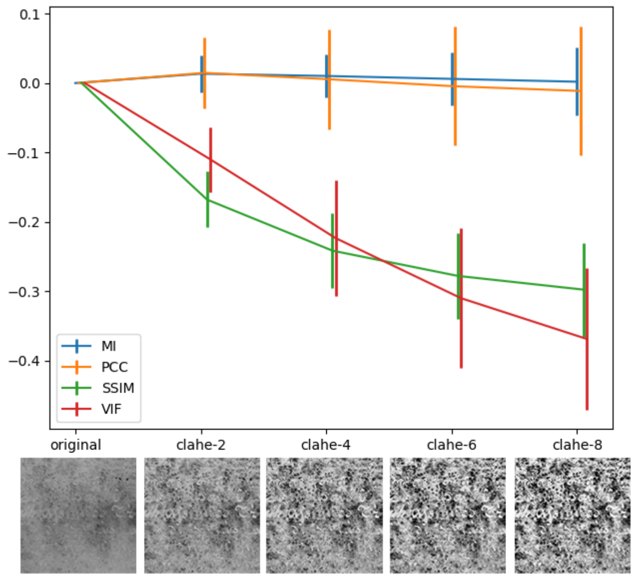
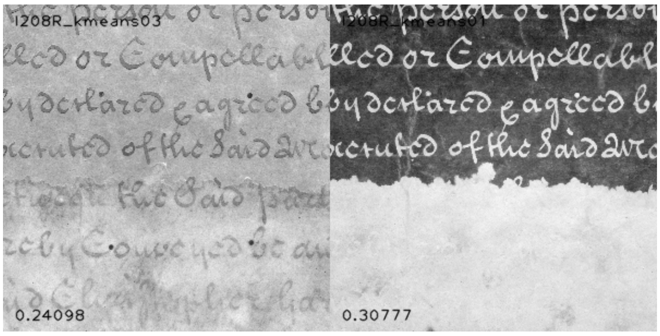
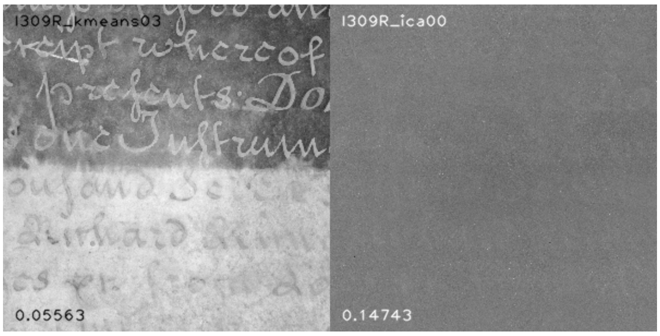


Fig. 2: The effect of applying CLAHE with increasing clip limit to the processed images before comparison with the respective metrics. Standard deviations are shown as vertical bars. The images below the plot give an example of a source image and resulting contrast-enhanced images. Note that the background structure is enhanced as well as the text.



(a) Irregularity in MI score



(b) Irregularity in PCC score

Fig. 3: Examples of wrong ratings. Images on the right were rated higher than images on the left.

found frequently. Figure 3 shows examples. The reason for those errors have not been investigated yet.

To definitely validate the feasibility of the approach, a user study is necessary to obtain a strong ground truth dataset containing subjective quality ratings from multiple individuals. Such datasets are the basis for any quantitative evaluation of image quality metrics, just as it is the case for general IQA problems [4], [28].

IV. Conclusion

In this paper we have surveyed the approach to assess the quality of readability enhancement methods by comparison with intact reference images, both theoretically and experimentally, and formulated it as a special case of Full-Reference Image Quality Assessment. Intuitively the approach is sensible, because the goal of any digital restoration is to produce results as similar to the originals as possible. Using four relatively simple image comparison metrics we produced visually convincing rankings of processed images; however, cases where the method fails were observed as well. In general, the four tested metrics correlate well, with Mutual Information and Pearson Correlation Coefficient showing the strongest agreement. We also showed that generic contrast enhancements have no positive effects on the comparison score and identified Mutual Information as the most stable metric in this regard. However, for a definite confirmation of the validity of this approach, a set of test images with expert-rated readability scores is required. To this end, a systematic user study is necessary. Only with this prerequisite can an improvement of the method be attempted, that is, the development of a of a more specialized and stable metric for image comparison. These attempts can also pave the way for the exploration of No-Reference IQA methods for readability assessment, which would be the optimal solution for this problem.

References

[1] S. A. Amirshahi, M. Pedersen, and S. X. Yu, "Image Quality Assessment by Comparing CNN Features between Images," *Journal of Imaging Science and Technology*, vol. 60, no. 6, pp. 60 410–1–60 410–10, 2016.

[2] C. T. C. Arsene, S. Church, and M. Dickinson, "High performance software in multidimensional reduction methods for image processing with application to ancient manuscripts," *Manuscript Cultures*, vol. 11, pp. 73–96, 2018.

[3] T. O. Aydn, K. I. Kim, K. Myszkowski, and H.-p. Seidel, "NoRM : No-Reference Image Quality Metric for Realistic Image Synthesis," *Computer Graphics Forum*, vol. 31, no. 2, 2012.

[4] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "On the use of deep learning for blind image quality assessment," *Signal, Image and Video Processing*, vol. 12, no. 2, pp. 355–362, 2018.

[5] S. Bosse, D. Maniry, K.-r. Müller, T. Wiegand, and W. Samek, "Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, 2018.

[6] S. Brenner, "On the Use of Artificially Degraded Manuscripts for Quality Assessment of Readability Enhancement Methods - Dataset & Code," DOI: 10.5281/zenodo.2650152FNo, 2019. [Online]. Available: https://doi.org/10.5281/zenodo.2650152

[7] R. L. Easton, W. A. Christens-Barry, and K. T. Knox, "Spectral image processing and analysis of the Archimedes Palimpsest," *European Signal Processing Conference*, no. Eusipco, pp. 1440–1444, 2011.

[8] S. Faigenbaum, B. Sober, A. Shaus, M. Moinester, E. Piasetzky, G. Bearman, M. Cordonsky, and I. Finkelstein, "Multispectral images of ostraca: Acquisition and analysis," *Journal of Archaeological Science*, vol. 39, no. 12, pp. 3581–3590, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.jas.2012.06.013

[9] A. Giacometti, A. Campagnolo, L. MacDonald, S. Mahony, S. Robson, T. Weyrich, and M. Terras, "UCL Multispectral Processed Images of Parchment Damage Dataset," DOI: 10.14324/000.ds.1469099, 2015. [Online]. Available: http://discovery.ucl.ac.uk/id/eprint/1469099

[10] A. Giacometti, A. Campagnolo, L. MacDonald, S. Mahony, S. Robson, T. Weyrich, M. Terras, and A. Gibson, "The value of critical destruction: Evaluating multispectral image processing methods for the analysis of primary historical texts," *Digital Scholarship in the Humanities*, vol. 32, no. 1, pp. 101–122, 2017.

[11] L. Glaser and D. Deckers, "The Basics of Fast-scanning XRF Element Mapping for Iron-gall Ink Palimpsests The Basics of Fast-scanning XRF Element Mapping for Iron-gall Ink Palimpsests," *Manuscript Cultures*, vol. 7, no. December 2013, 2016.

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, pp. 275–276, http://www.deeplearningbook.org.

[13] F. Hollaus, M. Diem, S. Fiel, F. Kleber, and R. Sablatnig, "Investigation of Ancient Manuscripts based on Multispectral Imaging," *DocEng 2015 - Proceedings of the 2015 ACM Symposium on Document Engineering*, no. 1, pp. 93–96, 2015.

[14] F. Hollaus, M. Diem, and R. Sablatnig, "Improving OCR accuracy by applying enhancement techniques on multispectral images," *Proceedings - International Conference on Pattern Recognition*, pp. 3080–3085, 2014.

[15] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1733–1740, 2014.

[16] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "elastix: A toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, Jan 2010.

[17] L. Likforman-Sulem, J. Darbon, and E. H. Smith, "Enhancement of historical printed document images by combining Total Variation regularization and Non-local Means filtering," *Image and Vision*

Computing, vol. 29, no. 5, pp. 351–363, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.imavis.2011.01.001

[18] K.-Y. Lin and G. Wang, "Hallucinated-IQA: No-Reference Image Quality Assessment via Adversarial Learning," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 732–741, 2018.

[19] L. MacDonald, A. Giacometti, A. Campagnolo, S. Robson, T. Weyrich, M. Terras, and A. Gibson, "Multispectral imaging of degraded parchment," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7786 LNCS, pp. 143–157, 2013.

[20] S. Mindermann, "Hyperspectral Imaging for Readability Enhancement of Historic Manuscripts Hyperspectral Imaging for Readability Enhancement of Historic Manuscripts," Master's thesis, TU München, 2018.

[21] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[22] A. K. Moorthy and A. C. Bovik, "Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

[23] E. Pouyet, S. Devine, T. Grafakos, R. Kieckhefer, J. Salvant, L. Smieska, A. Woll, A. Katsaggelos, O. Cossairt, and M. Walton, "Revealing the biography of a hidden medieval manuscript using synchrotron and conventional imaging techniques," *Analytica Chimica Acta*, vol. 982, pp. 20–30, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.aca.2017.06.016

[24] E. Salerno, A. Tonazzini, and L. Bedini, "Digital image analysis to enhance underwritten text in the Archimedes palimpsest," *International Journal on Document Analysis and Recognition*, vol. 9, no. 2-4, pp. 79–87, 2007.

[25] D. Shamonin, E. Bron, B. Lelieveldt, M. Smits, S. Klein, and M. Staring, "Fast parallel image registration on cpu and gpu for diagnostic classification of alzheimer's disease," *Frontiers in Neuroinformatics*, vol. 7, p. 50, 2014. [Online]. Available: https://www.frontiersin.org/article/10.3389/fninf.2013.00050

[26] A. Shaus, S. Faigenbaum-Golovin, B. Sober, and E. Turkel, "Potential Contrast – A New Image Quality Measure," *Electronic Imaging*, vol. 2017, no. 12, pp. 52–58, 2017.

[27] H. R. Sheikh and A. C. Bovik, "Image Information and Visual Quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[28] H. Sheikh, Z. Wang, L. Cormack, and A. Bovik, "LIVE Image Quality Assessment Database Release 2." [Online]. Available: http://live.ece.utexas.edu/research/quality

[29] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, Sep 1997. [Online]. Available: https://doi.org/10.1023/A:1007958904918

[30] Z. Wang, A. C. Bovik, H. R. Sheikh, S. Member, E. P. Simoncelli, and S. Member, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[31] D. Xin, L. Ma, S. Song, and A. Parameswaran, "How Developers Iterate on Machine Learning Workflows – A Survey of the Applied Machine Learning Literature," 2018.

[32] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., 2014, pp. 818–833.

[33] X. Zhu, S. Member, and P. Milanfar, "Automatic Parameter Selection for Denoising Algorithms Using a No-Reference Measure of Image Content," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3116–3132, 2010.