Semi-supervised Convolutional Triplet Neural Networks for Assessing Paper Texture Similarity

Leah Lackey, Arick Grootveld, and Andrew G. Klein
Electrical & Computer Engineering
Western Washington University
Bellingham, WA 98225
Email: {lackeyl, grootva, andy.klein}@wwu.edu

***Abstract*—In the context of papers used in the graphic arts, including silver gelatin, inkjet, and wove papers, prior work has studied measures of texture similarity for purposes of classifying such papers. The majority of prior work has been based on classical image processing approaches such as Fourier, wavelet, and fractal analysis. In this work, recent advances in deep learning are used to develop a texture similarity approach for measuring paper texture similarity. Since the available datasets generally lack labels, the convolutional neural network is trained using triplet loss to minimize the feature distance of tiles from the same image while simultaneously maximizing the feature distance of tiles drawn from different images. The approach is tested on three paper texture image databases considered in prior works and the results suggest the proposed approach achieves state-of-the-art performance.**

***Index Terms*—image texture analysis, machine learning, digital humanities**

I. Introduction

Surface texture is a critical, defining feature of paper used in the graphic arts as it impacts the visibility of fine detail. Texture analysis of paper used in the graphic arts provides important insights to the community of art investigators at museums and other art institutions, such as helping to validate authenticity, identifying purpose, and making important connections in the history of an artist or set of artists that may have worked together [1], [2]. An effort to address the issue of texture classification has been ongoing, starting with the Historic Photographic Paper Classification Challenge (HPPC) [1]. This effort has led to the creation of several datasets as well as numerous works that have proposed measures of texture similarity, including such approaches as multi-scale analysis (using anisotropic wavelets [2] or fractals [3]), non-semantic feature extraction (eigentextures [4], random-feature textons [1], deviation of local Gabor features [5]), local radius index [6], and restricted Boltzmann machines [7].

Advances in machine learning have raised the prospect of automated classification of paper in which the learning algorithm implicitly develops the classification features. It may be used not only to reinforce the classifications of human experts, but also to perhaps identify human classification errors. In this paper we explore the application of deep learning as a means to perform feature extraction for assessing the similarity of two textures. While machine learning has been used for clustering of art historical papers [8], [9], the prior texture similarity approaches [1]–[7] in this domain generally use more classical signal and image processing techniques, and this work contributes a distinctly new approach to the diverse toolbox of texture similarity methods.

A machine learning approach using so-called "triplet" neural networks [10] has shown success in the context of facial recognition for measuring the likeness of images of faces [11]. The triplet neural network approach is interesting for its ability to learn the features themselves. By simultaneously minimizing the distance between "like" images while maximizing the distance between "unlike" images, the triplet loss approach to training a neural netwok has shown promising results on a number of known datasets [12]. In this paper, we employ the triplet neural network approach for partially supervised learning on datasets of image textures, and explore the use of features extracted by the algorithm as a measure of texture similarity between pairs of texture images.

We subsequently conduct experiments on benchmark datasets to empirically validate the power of this neural network approach for feature learning and as a means for performing automated texture similarity assessment. We compare the performance of this approach with the more classical image-processing approaches that have published results on these same datasets [1]–[7].

II. Datasets

Since the available data for a particular application dictates to some degree what range of machine learning approaches is feasible (e.g., supervised or unsupervised), we begin with a discussion of the texture datasets of art historical papers. All the datasets used in this work consist of raking light photomicrographs of papers, acquired with a digital imager fitted with a zoom imaging lens. The field of view of the digital imager spans a physical area of 1.00×1.35 cm on the paper, and produces images with a resolution of 1536×2080 pixels. A 3-inch LED line light placed at a 25° raking angle to the surface of the papers illuminates the surface, and serves to enhance the highlights and shadows so that surface features are more clearly visible during image capture. For this reason, the raking light is used extensively in the examination of works

This work was supported by a Research Experiences for Undergraduates (REU) Supplement to National Science Foundation award 1836695 and a Jarvis Memorial Summer Research Award.

Fig. 1. Three types of similarity groups used to create labeled datasets.

of art, and produces images that can be used for automated texture classification.

Four datasets of photomicrographs illuminated with raking light were used as part of this work:

- **LML Silver Gelatin**. The Yale Lens Media Lab (LML) Reference Collection of Photographic Papers, perhaps the largest database of silver gelatin papers in the world, containing thousands of samples from 65 manufacturers and more than 360 brands. The dataset includes over 2,000 photomicrographs [1], though only a subset of 1,597 images used as part of [8] were used here.
- **HPPC Silver Gelatin**. A set of 120 images of silver gelatin paper, selected from the LML Reference Collection, and created as part of the HPPC. Ninety of the images in this dataset are from one of three similarity groups shown in Fig. 1. In addition, 30 sheets of interest to art conservators representing the diversity of silver gelatin photographic papers are included in the database. This dataset has been described in more detail in [13].
- **HPPC Inkjet**. A set of 120 images of inkjet paper, selected from the Henry Wilhelm Reference Collection, also created as part of the HPPC. Identical to the HPPC silver gelatin dataset just above, this dataset is comprised of 90 images from the same three similarity groups as well as 30 diversity samples. This dataset has been described in more detail in [14].
- **HPPC Wove**. A set of 180 images of wove paper, imaged from *Specimens* [15], a 1953 publication of the Stevens-Nelson Paper Corporation, and also created as part of the HPPC. The images in this dataset were obtained from both the front (recto) and back (verso) sides of the paper, thus there are 90 recto and 90 verso images. Furthermore, 120 of the images are from the first two similarity groups shown in Fig. 1, while 60 of the images are again included to represent the diversity of wove papers. This dataset has been described in more detail in [16].

One constraint in using these datasets with machine learning is that all of the prior texture similarity approaches reserved all 300 images from the similarity groups for testing algorithm performance. Thus, for a fair comparison any machine learning-based texture classification scheme would need at least those 300 images to be set aside into the test set, as shown in Table I. Note that this partitioning results in the maximally sized training set and minimally sized test set that permits a fair comparison with prior work.

While in total there are nearly two thousand images across

TABLE I
TRAIN/TEST DATASET PARTITION

dataset	Labeled?	# assigned to training set	# assigned to test set
LML Silver Gelatin	N	1477	0
HPPC Silver Gelatin	Y	30	90
HPPC Inkjet	Y	30	90
HPPC Wove	Y	60	120
Total	-	**1597**	**300**

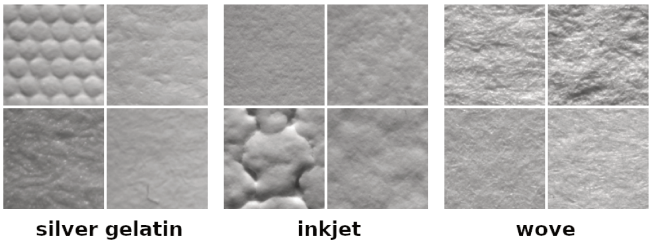

Fig. 2. Four example paper textures for each of the three types.

these four datasets which ought to be a sufficiently large collection of data to enable modern machine learning algorithms, there are two limiting factors evident from Table I. The first factor is that the much larger LML collection consists of *unlabeled* textures where texture affinities between all pairs of images have not been categorized by a domain expert, thus leaving just $30 + 30 + 60 = 120$ labeled images in the (maximally sized) training set which is likely to pose a challenge for traditional supervised learning approaches. The second factor is that, collectively, there is a large imbalance across the three types of paper (silver gelatin, inkjet, and wove). Since the manufacturing processes and even the manufacturing dates between these three categories of paper are so different [1], [14], [16], it is reasonable to expect differences in the surface features, and the presence of only 30 inkjet and 60 wove papers in the training set is likely to negatively impact performance. The examples in Fig. 2 provide visual evidence of these surface differences.

III. Technical Approach

A. Motivation for Triplet Loss

The prior approaches for assessing texture similarity of art historical papers [1]–[7] work by first proposing some mechanism to extract features from each image (e.g., wavelets, Fourier bases, fractals) and the features are subsequently used to compute pairwise distances between all images in the dataset. We adopt the same basic operations consisting of feature extraction followed by distance computation. Where our approach differs significantly from the majority of prior, more classical image processing approaches, however, is in the feature extraction step: we employ a data-driven approach where we attempt to *learn* an appropriate feature extraction method (i.e., a mapping from the input image to an embedding).

The idea of training a neural network to minimize so-called *triplet loss* has been proposed [10], [11], and led to

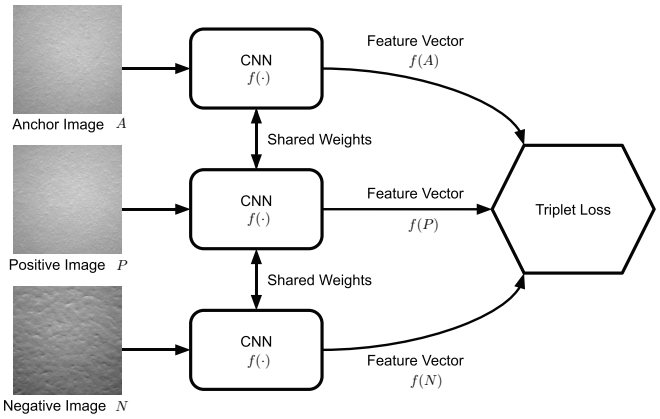

Fig. 3. Training a triplet neural network amounts to finding a function $f(\cdot)$ that minimizes the Euclidean distance between $f(A)$ and $f(P)$ while maximizing the Euclidean distance between $f(A)$ and $f(N)$.

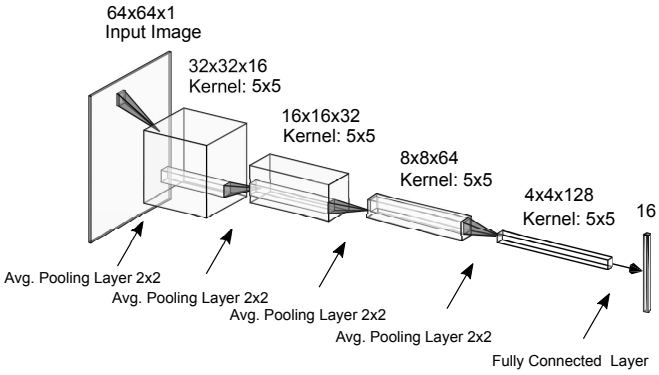

Fig. 4. Convolutional neural network architecture.

great improvements in the domain of facial recognition. The approach trains a neural network to perform feature extraction in a way that minimizes the distance between "matches" while maximizing the distances between "non-matches". Let us assume that the neural network accepts as input an image A and produces at its output a feature vector $f(A)$ of reduced dimension. The network is trained by forming "triplets" consisting of three inputs: an *anchor* image A and *positive* image P which are known to be matches, and additional image N called the *negative* which is known to not be a match to either the anchor or the positive. The neural network $f(\cdot)$ is then trained to minimize the loss given by

$$\mathcal{L}(A,P,N) = \max\left\{||f(A) - f(P)||_2^2 - ||f(A) - f(N)||_2^2 + \alpha, 0\right\}$$

where the parameter α is a constant (sometimes referred to as the "margin") added to avoid the trivial case where the network outputs the same feature vector for all inputs. Note that the term $||f(A)-f(P)||_2^2$ is simply the squared Euclidean distance between anchor and positive feature vectors (which we seek to minimize), while $||f(A)-f(N)||_2^2$ is the squared Euclidean distance between anchor and negative feature vectors (which we seek to maximize). An overview of the approach is shown in Fig. 3. Because minimizing triplet loss results in an end-to-end learning between the input image and distances in the feature vector space, the approach directly optimizes the neural network for the final task (i.e., computing distances between images).

As described above in Section II, the available datasets do not offer a training set that contains sufficient labeled data to permit the use of conventional supervised learning approaches, and the texture affinities within the training set are unknown. Because the triplet neural network approach requires knowledge of matches and non-matches to appropriately select the anchor, positive, and negative images, it appears at first glance that the unlabeled training data precludes the use of triplet loss as a feasible training mechanism. To get around this issue and permit the use of triplet loss in a partially-supervised fashion, we do two things:

1) We split the images into 16 tiles and use the tiles as inputs to the neural network so that we can indeed form triplets containing known matches.
2) We make the assumption that tiles from two different images in the training set are *not* identical textures.

This assumption is almost certainly false on occasion because some of the 1,477 unlabeled images in the training set are likely to be identical textures, or nearly so. However, this assumption allows us to create triplets by always selecting the anchor A and positive P to be tiles from the same image, while always selecting the negative N to be a tile from a different image (presumed, perhaps incorrectly, to be a distinct texture).

B. Convolutional Neural Network Details

Having motivated the use of triplet loss combined with tiling of the images as the core idea of our partially supervised technical approach, we now describe the details of the underlying convolutional neural network (CNN). Through experimentation we have shown that the choice of CNN architecture used is perhaps not so important, as multiple architectures yielded encouraging results. As such, we adopted a fairly standard CNN architecture shown in Fig. 4.

We preprocess the raw images in the dataset using the following five steps: (i) extract a 1024×1024 pixel snippet from the middle of each image to avoid the impact of vignetting, (ii) convert the image to greyscale to focus on surface texture rather than color, (iii) downsample the image by a factor of 4 to yield a 256×256 pixel image due to prior work which suggested that the smallest scales are less helpful in classifying texture [8], (iv) normalize all images to consist of pixel values between 0 and 1, and (v) perform a 4×4 tiling of the images which results in 16 tiles of size 64×64 pixels. In the training stage, we use data augmentation by randomly sampling an additional 16 tiles from each image and applying random adjustments to brightness and contrast. Data augmentation is known to improve the ability of a network to generalize, and the adjustments to brightness and contrast are intended to minimize the impact of camera exposure in the training of the CNN. The augmented training set then consists of 32 tiles per image, or 51,104 tiles. Three randomly selected tiles from

each image are set aside in a validation set to be used as part of training.

As shown in Fig. 4, the preprocessed 64×64 pixel tiles are input into the CNN which consists of 4 convolutional layers with ReLU activation functions and 2×2 average pooling between each layer, followed by a final fully connected layer at the output which produces a length 16 feature vector. The feature vector is subsequently ℓ^2-normalized to a unit hypersphere. The resulting CNN consists of 171,152 trainable parameters. During training a batch size of 512 randomly selected triplets was used. The tiles reserved as part of the validation set are used to compute the mean reciprocal rank (discussed below in Section IV), and this metric is used to determine when to halt training. The TensorFlow implementation is available at [17], and takes approximately 2 hours to train on dual NVIDIA GTX 1080 GPUs.

While the CNN operates on tiles rather than whole images, we ultimately desire to know the distance between two entire images rather than the constituent tiles. As such, we compute the feature vector for an entire image by taking the centroid of its 16 tile feature vectors and then scaling by one half. That is, if A_{ij} represents the 16 tiles of image A for $i, j \in \{1, 2, 3, 4\}$, then the image feature vector v is computed from the tile feature vectors $f(A_{ij})$ via $v = \frac{1}{2}\left(\frac{1}{16}\sum_{i,j} f(A_{ij})\right)$, where the unit-normalization of $f(A_{ij})$ discussed previously guarantees that $||v|| \leq \frac{1}{2}$. The distance between any two images, then, is computed by taking the Euclidean distance of their corresponding image feature vectors. The factor of one half is included so that the Euclidean distance between any two image feature vectors v_1 and v_2 satisfies $||v_1 - v_2|| \leq ||v_1|| + ||v_2|| \leq 1$ by the triangle inequality. That is, distances are always between 0 and 1.

IV. Results

A. Quantitative Results

To quantify similarity assessment performance, we adopt metrics used in the information retrieval community to assess the performance of our approach compared to prior approaches. As in the prior work, here we consider performance on the test set for each of the three types of paper independently. The metrics employed are based not on the distances themselves, but on the *rank* of true matches when, for a given query image, all other images in the dataset are ordered by increasing distance to the query image. In particular, we consider three performance metrics: (i) *precision at one* (P@1) which is the mean fraction of time that the top ranked match (having smallest distance to the query image) is a true match, (ii) *mean reciprocal rank* (MRR) which measures the mean inverse rank of the first true match [18], and (iii) *mean average precision* (MAP) [18]. The MAP is calculated as follows: for each query image and positive integer n less than or equal to the size of the dataset, compute the fraction of the n highest ranked images that are true matches, and then average these fractions over all values of n for which the nth highest ranked image was actually a true match; then, average these values across all images. The compared performance metrics to the top-performing prior works are reported for silver gelatin paper in Table II, for inkjet paper in Table III, and for wove paper in Table IV. Results for the prior approaches were reported in [6], [16].

TABLE II
SILVER GELATIN RETRIEVAL MEASURES

Algorithm	P@1	MRR	MAP
Triplet NN	96.7%	97.7%	87.6%
LRI	98.9%	99.1%	91.5%
HWT	62.2%	76.9%	65.0%
PASFA	85.6%	89.9%	73.1%

These results provide compelling evidence that the triplet CNN-based approach is among the top performing approaches, and in some cases (such as the recto side of the wove paper), is the top-performing approach. Given that only a small fraction of the training set consisted of wove paper, it is somewhat surprising that the algorithm is able to perform so well on the wove papers.

B. Qualitative Results

Figure 5 visualizes the distances computed between each pair samples in the silver gelatin test set and compares them to the expert assessment. Here, dark shades represent small distances whereas light shades represent large distances. Figure 6 shows the same comparison for the case of inkjet images. The results for wove paper have been omitted due to lack of space. The 10×10 black squares along the main diagonal correspond to 10 "like" images from one of the three similarity groups depicted in Fig. 1; the presence of these black squares in both the expert assessment as well as the proposed approach provides visual evidence of very low distances amongst the different samples from the same sheet, same package, or same manufacturer designation.

Overall, this qualitative picture provides convincing evidence of the feasibility of our proposed technique in performing automated texture similarity assessment. While our proposed technique is very different from prior techniques, we note that in some cases our method yields results that disagree with the expert assessment but agree with prior automated approaches. For example, in the right side of Fig. 6, our results suggest that the third sample in the inkjet test set is somewhat of an outlier as evidenced by the grey band passing through the third row and third column. While this is not present in the expert assessment, it was indeed evident in the results of several of the prior works reported in [14].

V. Conclusions and Acknowledgments

In this paper, we presented a triplet neural network-based approach for performing automated texture similarity assessment. The results are encouraging and demonstrate the promise of this approach as a useful tool for texture classification; however, rather than focusing on the minute performance differences with respect to prior work we emphasize that automated assessment of texture similarity can be achieved

TABLE III
INKJET PAPER RETRIEVAL MEASURES

Algorithm	P@1	MRR	MAP
Triplet NN	86.7%	90.6%	77.3%
LRI	90.0%	92.9%	79.8%
HWT	87.8%	90.4%	77.6%
PASFA	87.8%	90.8%	77.7%

TABLE IV
WOVE PAPER RETRIEVAL MEASURES

Algorithm	Recto			Verso		
	P@1	MRR	MAP	P@1	MRR	MAP
Triplet NN	100.0%	100.0%	91.4%	100.0%	100.0%	93.6%
LRI	95.0%	97.5%	94.4%	100.0%	100.0%	96.8%
HWT	98.3%	99.2%	95.1%	100.0%	100.0%	97.7%
PASFA	95.0%	96.9%	73.6%	88.3%	92.2%	65.7%

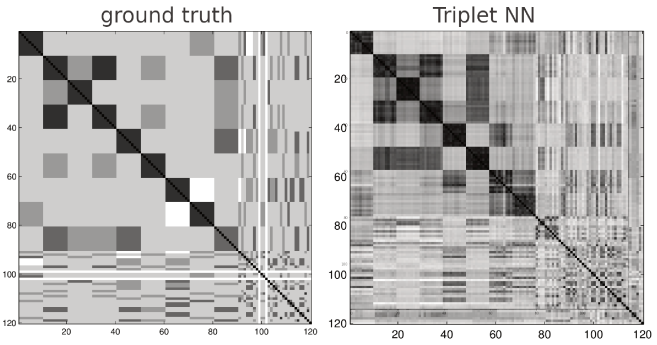

Fig. 5. Silver gelatin distances, with expert assessment (left) compared to the distance matrices computed using our proposed approach. Distances range from black (smallest distance) to white (largest distance).

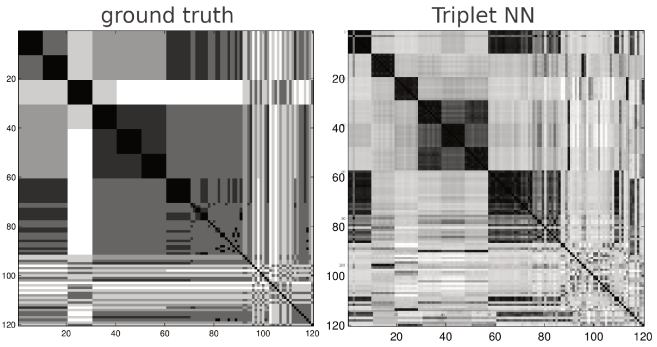

Fig. 6. Inkjet distances, with expert assessment (left) compared to the distance matrices computed using our proposed approach. Distances range from black (smallest distance) to white (largest distance).

from tools very different in principle, and this work here provides yet one more for the toolbox of approaches. Future work could investigate combining this approach with other, prior work, or of incorporating this approach in a user-friendly software tool that would broaden its use by domain experts. In addition, future work might investigate texture affinities in existing datasets focused on specific artists, such as F. Holland Day [19], Moholy, Matisse, or others. Finally, the authors wish to acknowledge Paul Messier (Institute for the Preservation of Cultural Heritage at Yale University), Peggy Ellis (Institute of Fine Arts at NYU), and Henry Wilhelm (Wilhelm Imaging Research) for the use of images and data that were instrumental to this work.

REFERENCES

[1] C. R. Johnson, P. Messier, W. A. Sethares, A. G. Klein, C. Brown, A. H. Do, P. Klausmeyer, P. Abry, S. Jaffard, H. Wendt *et al.*, "Pursuing automated classification of historic photographic papers from raking light images," *Journal of the American Institute for Conservation*, vol. 53, no. 3, pp. 159–170, 2014.

[2] P. Abry, S. G. Roux, H. Wendt, P. Messier, A. G. Klein, N. Tremblay, P. Borgnat, S. Jaffard, B. Vedel, J. Coddington *et al.*, "Multiscale anisotropic texture analysis and classification of photographic prints: Art scholarship meets image processing algorithms," *IEEE Signal Processing Magazine*, vol. 32, no. 4, pp. 18–27, 2015.

[3] A. G. Klein, A. H. Do, C. A. Brown, and P. Klausmeyer, "Texture classification via area-scale analysis of raking light images," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov. 2014, pp. 1114–1118.

[4] W. A. Sethares, A. Ingle, T. Krč, and S. Wood, "Eigentextures: An SVD approach to automated paper classification," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov 2014, pp. 1109–1113.

[5] D. Picard and I. Fijalkow, "Second order model deviations of local Gabor features for texture classification," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov 2014, pp. 917–920.

[6] Y. Zhai and D. L. Neuhoff, "Photographic paper classification via local radius index metric," in *Image Processing (ICIP), 2015 IEEE International Conference on*, Sept 2015, pp. 1439–1443.

[7] A. Sangari and W. Sethares, "Paper texture classification via multi-scale restricted Boltzman machines," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov 2014, pp. 482–486.

[8] A. G. Klein, P. Messier, A. L. Frost, D. Palzer, and S. L. Wood, "Deep learning classification of photographic paper based on clustering by domain experts," in *2016 50th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2016, pp. 139–143.

[9] K. R. Basinet, A. G. Klein, P. Abry, S. Roux, H. Wendt, and P. Messier, "Performance of two multiscale texture algorithms in classifying silver gelatin paper via k-nearest neighbors," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 1028–1032.

[10] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification." *Journal of Machine Learning Research*, vol. 10, no. 2, 2009.

[11] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. on computer vision and pattern recognition*, 2015, pp. 815–823.

[12] A. Hermans*, L. Beyer*, and B. Leibe, "In Defense of the Triplet Loss for Person Re-Identification," *arXiv preprint arXiv:1703.07737*, 2017.

[13] P. Messier and C. R. Johnson, "Automated surface texture classification of photographic print media," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov. 2014, pp. 1105–1108.

[14] P. Messier, C. R. Johnson, H. Wilhelm, W. A. Sethares, A. G. Klein, P. Abry *et al.*, "Automated surface texture classification of inkjet and photographic media," in *Proc. Intl. Conf. on Digital Printing Technologies (NIP 29)*, Sep. 2013.

[15] *Specimens: A Stevens-Nelson Paper Catalogue*. New York: Stevens-Nelson Paper Corporation, 1953.

[16] P. Abry, A. G. Klein, P. Messier, S. Roux, M. H. Ellis, W. A. Sethares, D. Picard, Y. Zhai, D. L. Neuhoff, H. Wendt *et al.*, "Wove paper analysis through texture similarities," in *2016 50th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2016, pp. 144–148.

[17] L. Lackey, A. Grootveld, and A. G. Klein. (2020) Source code for paper texture classification using a triplet convolutional neural network. [Online]. Available: https://github.com/aspectlab/tripletpapertexture

[18] E. M. Voorhees, "Variations in relevance judgments and the measurement of retrieval effectiveness," *Inf. Process. Manage.*, vol. 36, no. 5, pp. 697–716, Sep. 2000.

[19] P. Abry, S. Roux, A. Lundgren, P. Messier, A. Klein, H. Wendt, and S. Jaffard, "F. Holland Day art photographic paper clustering: Automated procedures to assist photograph conservators," in *Proc. of the 2019 GRETSI Conf.*, Aug. 2019.