Assessing Paper Texture Similarity in Matisse Lithographs Using a Triplet Neural Network

Karen M. Aguilar, Shelby Powers, Leah Lackey, Arick Grootveld, and Andrew G. Klein
Electrical and Computer Engineering
Western Washington University
Bellingham, WA 98225
Email: {aguilak2, powerss8, lackeyl, grootva, andy.klein}@wwu.edu

***Abstract*—This work explores the use of a triplet neural network for assessing the similarity of paper textures in a collection of Henri Matisse's lithographs. The available dataset contains digital photomicrographs of papers in the lithograph collection, consisting of four views: two raking light orientations and both sides of the paper. A triplet neural network is first trained to extract features sensitive to anisotropy, and subsequently used to ensure that all papers in the dataset are in the same orientation and side. Another triplet neural network is then used to extract the texture features that are used to assess paper texture similarity. These results can then be used by art conservators and historians to answer questions of art historical significance, such as artist intent.**

I. Introduction

The last decade has seen growing interest in the use of signal/image processing and machine learning to help answer research questions in cultural heritage, including those in art scholarship and preservation [1]–[3]. The increasing availability of vast datasets combined with the rapid advance of machine learning techniques has served, in part, to fuel this cross-disciplinary research area, and has permitted the community of researchers to ask questions which were previously impossible to answer. In this work, we apply recent machine learning techniques to a dataset of 215 lithographs by the French artist Henri Matisse with the goal of identifying paper texture similarities or clusters within this dataset.

Texture analysis of papers of art historical interest has been studied extensively (see for example [4]–[6]), and a wide range of signal and image processing techniques have been applied to this problem, such as hyperbolic wavelet transforms [7], fractals [8], local radius index [9], and restricted Boltzmann machines [10], among others. Recently, a machine-learning-based approach to paper texture analysis that uses a triplet neural network was developed and shown to yield promising results in this domain when applied to several test datasets comprised of multiple paper types, including wove, silver gelatin, and inkjet papers [11].

In this paper, we review the triplet neural network approach that we developed previously in [11], and we adapt this approach to address the unique challenges posed by this dataset of Matisse lithographs. All of the prior datasets on which the triplet neural network approach was tested were relatively controlled test sets created from reference collections of example papers, and all papers were imaged with the same orientation and consistent front/back side as much as possible. In this dataset consisting of paper textures from an actual art collection, the orientation of the papers varies considerably, as the orientations may have varied purposefully by the artist, or may have varied due to inconsistencies in the print studio. Thus, before assessing texture similarity of the textures, the multiple images (orientation and paper side) of each paper texture need to be considered for possible permutations to achieve a consistent orientation of all the paper textures in the dataset. We develop and report on an approach that adapts the training process of the triplet neural network to extract features that can be used to permute the images as needed. Once the images in the dataset are appropriately permuted to have consistent orientation, we apply the texture feature extraction approach in [11] and then subsequently assess similarity between textures in the Matisse lithograph dataset.

Prior work that first analyzed this dataset [12] has used the Hyperbolic Wavelet Transform, and also gave attention to the problem of permuting images in the dataset to achieve consistent orientation. We compare the results generated by our approach to those reported by the authors of [12]. Our approach is quite different in that we use a triplet neural network; however, it is very encouraging that the two very different approaches agree on a large number of permutations and affinities within the dataset.

II. Matisse Lithograph Dataset

The images in the dataset consist of digital photomicrographs of 215 of Matisse's lithographs, and are from the collection of The Pierre and Tana Matisse Foundation. The prints date from 1925 through 1951, and the vast majority are printed on papers with an *Arches* watermark. While [12] provides more details of the dataset including example images, the papers are primarily wove papers and have attributes typical of high-quality printing papers made during the 20th century. The 215 prints are organized across 50 editions, where an edition is a group of numbered prints made at the same time; usually, but not always, prints within an edition are made on the same paper type.

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

978-1-6654-0595-9/21/$31.00 ©2021 IEEE

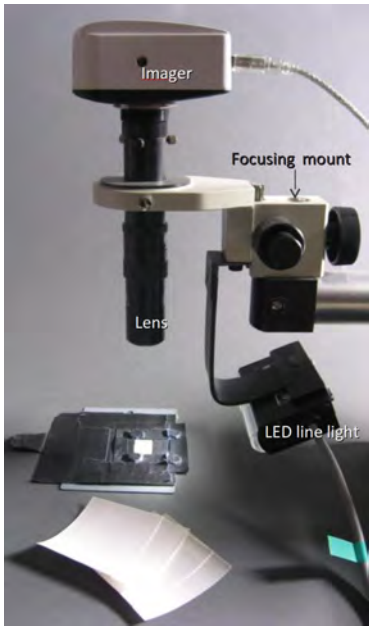


Fig. 1. Digital imager and raking light

In wove papermaking, the two sides of paper are referred to as the *screen* and *felt* sides, referring to the wire screen used to drain water from the pulp, and the felt on which the pulp is subsequently placed to dry. For paper made in this way, it is common for the screen and felt sides to exhibit different textures; as such, all papers were imaged on both the front (or *recto*) and back (or *verso*) sides.

The photomicrographs were acquired with a digital imager fitted with a zoom imaging lens, as shown in Fig. 1. The field of view of the digital imager spans a physical area of 1.0 cm × 1.2 cm on the paper, and produces images with a resolution of 2048×2448 pixels. A 3-inch LED line light placed at a 25° raking angle to the surface of the papers illuminates the surface and also enhances the highlights and shadows so that surface features are more clearly visible during image capture. Due to the possibility that these papers may exhibit anisotropy, the papers were imaged with the raking light oriented in two positions relative to the printed image: (i) parallel to the top of the print, and (ii) parallel to the side edge of the print. This led to four views for each paper, and we adopt the convention of referring to them as Recto-Top (RT), Recto-Side (RS), Verso-Top (VT) and Verso-Side (VS).

Finally, we note that cropped, downsampled, greyscale versions of these source images are used throughout this work, except occasionally in some of the figures. Cropping minimizes the impact of vignetting and lens distortion, downsampling is performed since the image resolution is higher than needed for performing texture analysis, and greyscale images prevent the machine learning algorithms from relying too heavily on color rather than texture. After performing these steps, the source images are transformed into 256×256 pixel greyscale images, which are subsequently tiled.

III. Triplet Neural Network

Here, we review the triplet neural network structure previously developed for paper texture classification in [11]. This approach uses a relatively traditional convolutional neural network (CNN) to perform *feature extraction*, thus mapping each image to a length-16 feature vector containing the relevant texture information. Subsequently, the feature vectors are used to arrive at the "distance" or dissimilarity of any pair of images; this is accomplished by computing the Euclidean distance between the two feature vectors. In this work, we use this same neural network structure for two different tasks, each with a different set of weights/kernels: (i) one whose weights we train to extract features sensitive to anisotropy, and (ii) a structurally identical CNN with weights trained as in [11], thus leading to a CNN that extracts features suitable for texture classification.

To train the CNN, we use a loss function called *triplet loss* which has been proposed in [13] for use in facial recognition applications. The approach trains a neural network to perform feature extraction in a way that minimizes the distance between "like" textures while maximizing the distances between "unlike" textures. 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 identical textures, and an additional image N called the *negative* which is known to be a different texture from 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 added to avoid the trivial case where the CNN outputs the same feature vector for all inputs. The term $||f(A)-f(P)||_2^2$ is 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. A graphical overview of the approach is shown in Fig. 2. 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).

The CNN structure we employ consists of four 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 that is ℓ^2-normalized to a unit hypersphere. Additional details of the various hyperparameters can be found in [11] or by reviewing the source code available in [14].

IV. Methodology and Results

In this section we describe how we train and use the two aforementioned CNN's to first permute the images in the dataset to have consistent orientation, and subsequently to

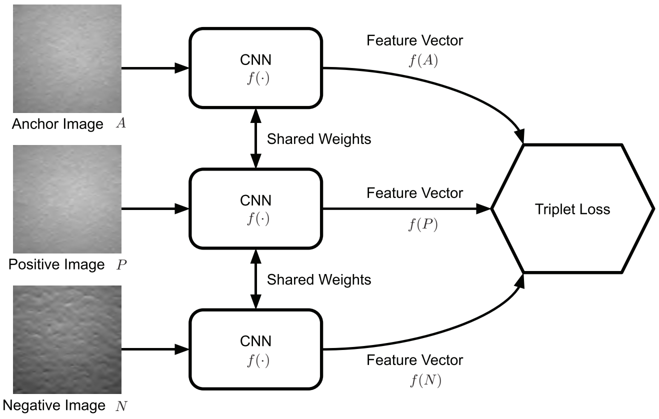


Fig. 2. Training a triplet neural network amounts to finding a function $f(\cdot)$ that minimizes the Euclidean distance between $f(A)$ and $f(P)$ while maximizing the Euclidean distance between $f(A)$ and $f(N)$.

extract texture features, compute pairwise distances between textures, and perform clustering to identify "like" textures within the dataset. Along the way, we provide results specific to the Matisse dataset and compare with prior work. Below, we refer to the CNN which is used to extract features sensitive to anisotropy as CNN$_{permute}$, while the CNN used to extract texture features is denoted CNN$_{texture}$. Again, we note that the structure of CNN$_{permute}$ and CNN$_{texture}$ is identical, and it is only the values of the weights which differ due to a different training process for each as described below.

A. Training CNN$_{permute}$

Since texture affinities within most art historical datasets are not known *a priori* and thus the data is generally unlabeled, and since the selection of the anchor, positive, and negative triplets requires labeled data, we make use of a tiling trick to train the neural network. Recall that a texture in the dataset consists of 4 images corresponding to the 4 views RT, RS, VT, and VS. We first split each 256×256 image in the dataset into 16 tiles of size 64×64 pixels. All tiles from the same image are of course identical textures, so we always draw the *anchor* and *positive* portions of the triplet as distinct tiles from the same image. To encourage the CNN to extract features sensitive to anisotropy, we draw the *negative* tile from the same paper, but from one of the other 3 views. Thus, the loss function implicitly selects features such that the distance between feature vectors from different views are maximized. At the same time, the loss function selects features such that the distance between two tiles from the same view of a single texture is minimized. In training this neural network, all images in the Matisse dataset are used, and before training begins the weights are pre-initialized to the values of CNN$_{texture}$ described below. Training of CNN$_{permute}$ is stopped after 400 epochs, using a batch size of 512 triplets.

B. Permuting images within editions

TABLE I
VIEW PERMUTATIONS

Notation	Description	Order
$\mathcal{A}$	original order	RT, RS, VT, VS
$\mathcal{B}$	recto/verso flip	VT, VS, RT, RS
$\mathcal{C}$	top/side rotation	RS, RT, VS, VT
$\mathcal{D}$	recto/verso and top/side	VS, VT, RS, RT

Next, we describe a procedure that uses the features computed by CNN$_{permute}$ to permute the images of a given paper so that all papers have the same orientation. For each of the 215 papers in the dataset and each of the 4 views per paper – giving 860 total textures – we use CNN$_{permute}$ to compute the length-16 feature vector for each texture, and subsequently we compute the Euclidean distance between all pairs of feature vectors within an edition. There are only four possible permutations of the four views that need to be considered, shown in Table I. Let $\pi_i \in \{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}\}$ denote a permutation for the ith paper, let $v_i \in \{\mathrm{RT}, \mathrm{RS}, \mathrm{VT}, \mathrm{VS}\}$ denote one of the four views of the ith paper, and let $\tilde{D}_{ij}(v_i, v_j)$ denote the Euclidean distance between the feature vectors corresponding to view v_i of paper i and view v_j of paper j. Let $D_{ij}(\pi_i, \pi_j)$ denote the summed Euclidean distance between each of the 4 pairs of feature vectors for matched views of the ith and jth papers under permutations π_i and π_j. For example, if $\pi_i = \pi_j = \mathcal{A}$, then

$$\begin{aligned} D_{ij}(\pi_i, \pi_j)\big|_{\pi_i=\pi_j=\mathcal{A}} &= \tilde{D}_{ij}(RT, RT) + \tilde{D}_{ij}(RS, RS) \\ &\quad + \tilde{D}_{ij}(VT, VT) + \tilde{D}_{ij}(VS, VS). \end{aligned}$$

Note that the right hand side of this expression is also the value of $D_{ij}(\pi_i, \pi_j)$ more generally whenever $\pi_i = \pi_j$ due to the associativity of addition. As another example, if $\pi_i = \mathcal{A}$ but $\pi_j = \mathcal{B}$ so that image i has the original ordering but image j has a recto/verso flip, then

$$\begin{aligned} D_{ij}(\pi_i, \pi_j)\big|_{\pi_i=\mathcal{A}, \pi_j=\mathcal{B}} &= \tilde{D}_{ij}(RT, VT) + \tilde{D}_{ij}(RS, VS) \\ &\quad + \tilde{D}_{ij}(VT, RT) + \tilde{D}_{ij}(VS, RS). \end{aligned}$$

Let $(\pi_1, \pi_2, \ldots, \pi_N)$ be an N-tuple of permutations corresponding to the N papers within an edition. Since there are four possible choices of π_i for each i, there are 4^N possible permutations of the N papers, though without loss of generality we could fix $\pi_1 = \mathcal{A}$, resulting in 4^{N-1} possible permutations. To find the best set of permutations π_i^* over a given edition, we use an exhaustive search to pick all the π_i to minimize the sum distance:

$$(\pi_1^*, \pi_2^*, \ldots, \pi_N^*) = \arg \min_{(\pi_1, \pi_2, \ldots, \pi_N)} \sum_{i,j} D_{ij}(\pi_i, \pi_j)$$

where the sum is over all N images in a given edition. The rationale for this choice is that papers, when permuted to be in the same orientation, ought to have the smallest possible sum distance over all matched views. That is, the "correct" permutation π_i^*, π_j^* ought to satisfy $D_{ij}(\pi_i^*, \pi_j^*) \leq D_{ij}(\pi_i, \pi_j)$ for all possible π_i, π_j.

While an exhaustive search can be prohibitively time consuming if N is large, we note that of the 50 editions in the Matisse dataset, the largest edition contains 10 papers, and thus there are at most a modest $4^9 \approx 2.6 \times 10^5$ permutations to search over for each of the 50 editions. Moreover, while this approach is more computationally expensive than the voting approach proposed in [12], it avoids the problem of having to break ties.

The result of applying this procedure was that permutations were needed across the following 14 of 49 editions in the dataset: 1249*, 1256, 1300, 1301, 1313*, 1324, 1327, 1367, 1403, 2589, 2592*, 2600, 2603*, 2607*. In most all cases, only a single paper needed to be permuted, and the vast majority of permutations were recto/verso flips (i.e., permutation $\mathcal{B}$). Moreover, half of the papers that needed permutations were designated as *essai* or *epreuve* prints which is encouraging as it is suspected that these are cases involving experimentation with different papers, orientations, and sides. The editions above with asterisks coincide with permutations identified in [12]. While our results suggest more flips than were identified by the authors of [12], we note that the list here includes all but one of those found in the prior work. Finally, an example of papers from edition 1313 where a permutation was required is shown in Fig. 3. There, it can be seen that permuting the second row leads to more consistent textures within columns.

C. Permuting images across editions

Having permuted the papers within editions to have the same orientation, we now address the problem of inter-edition permutations. First, we represent each edition with the centroid of all permuted feature vectors within the edition. Because there are 50 editions, we cannot directly adopt the exhaustive search used above since we would need to test $4^{49} \approx 3 \times 10^{29}$ permutations. Thus, we adopt a hierarchical approach by partitioning the 50 editions into 5 groups of 10, and applying the exhaustive search described above to resolve permutations within the partitions. Then, we repeat the process for all 5 groups by replacing the 10 images in each group with the centroid of its permuted feature vectors, and finally we resolve permutations across the 5 groups.

The results indicate that there are a large number of inter-edition permutations identified, much more than the relatively small number of intra-edition permutations above. Indeed, the permutations suggested by this approach resulted in a roughly equal number of the four permutation types shown in Table I, and thus roughly 75% of the editions needed a permutation that was different from the original ordering. It is reasonable to expect that intra-edition orientations would be fairly consistent since prints within an edition were generally made all at once. Meanwhile, across editions, there were much larger breaks in the workflow, with a new lithographic stone for each edition, and thus it is reasonable to expect that the papers between editions could differ significantly, both in the type of paper as well as the orientation and side. An example of several papers, each a representative from distinct editions, is shown in Fig. 4 in original and permuted views. Even though the rows are likely different papers with different textures, permuting rows again leads to more consistent textures within columns, suggesting that this is a compelling approach.

D. Using CNN$_{texture}$ to extract features, computing pairwise distances between all textures, and performing clustering

With all 215 papers in the dataset now permuted to have a consistent orientation across all 4 views, we finally extract texture features using the second triplet neural network, denoted CNN$_{texture}$, which was developed in [11]. Again, this neural network has the same structure as CNN$_{permute}$, though the weights are different since the training process is also different. We refer the reader to [11] for the details. Next, we use the texture features computed by CNN$_{texture}$ to compute the pairwise distances between all papers in the dataset, and subsequently we cluster the papers using k-medoids clustering with 18 clusters. A visual inspection of the clusters confirms that the approach yields very compelling results; however, deeper analysis and validation of these clusters requires collaboration with a domain expert in the area of art conservation.

V. Conclusion

Previously, the triplet neural network approach to paper texture classification was validated as a useful tool for assessing paper texture similarity on controlled datasets across a wide range of paper textures, including wove, silver gelatin, and inkjet papers. Here, we have provided evidence confirming that the approach is also suitable for use on a dataset of a single artist, where the texture differences between papers are likely to be more nuanced. Moreover, the approach was shown to be a useful technique for reorienting images and determining paper side. We have shown our results to domain experts in the field of art conservation, and they have found the results to be very compelling.

Future work will involve collaboration with experts in the domain of paper conservation and art history to provide a rigorous validation of these results as well as asking yet deeper questions about such issues as Matisse's preferences and artistic intent. The source code to produce these results was written primarily in TensorFlow, and is freely available at [14].

Acknowledgments

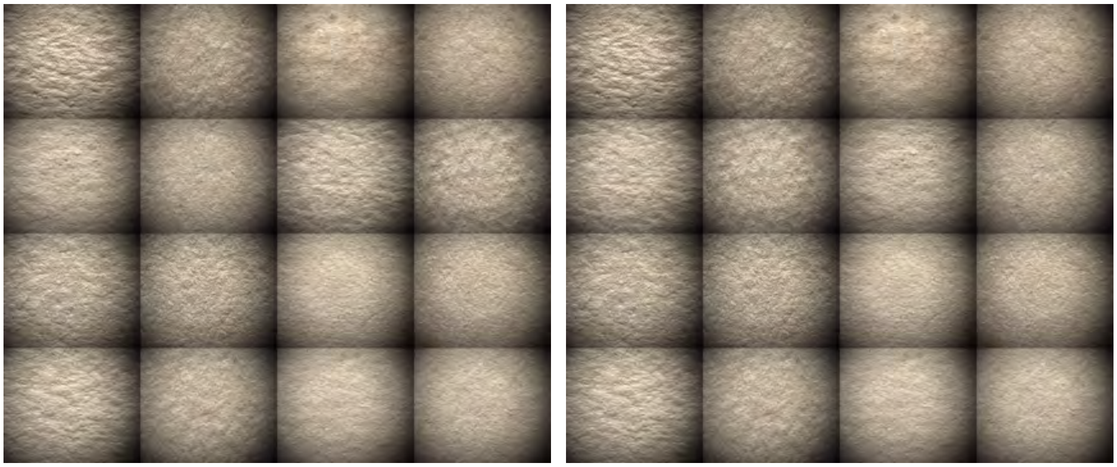
The authors would like to thank Paul Messier and his team in the Institute for the Preservation of Cultural Heritage at Yale University, Peggy Ellis at the Institute of Fine Arts NYU, Patrice Abry at ENS Lyon / CNRS, and The Pierre and Tana Matisse Foundation for the use of images and data that were instrumental to this work.

References

[1] *IEEE Signal Processing Magazine (special issue on "Recent Advances in Applications to Visual Cultural Heritage")*, vol. 22, no. 4, Jul. 2008.

[2] *Signal Processing (special issue on Image Processing for Digital Art)*, vol. 93, Mar. 2013.

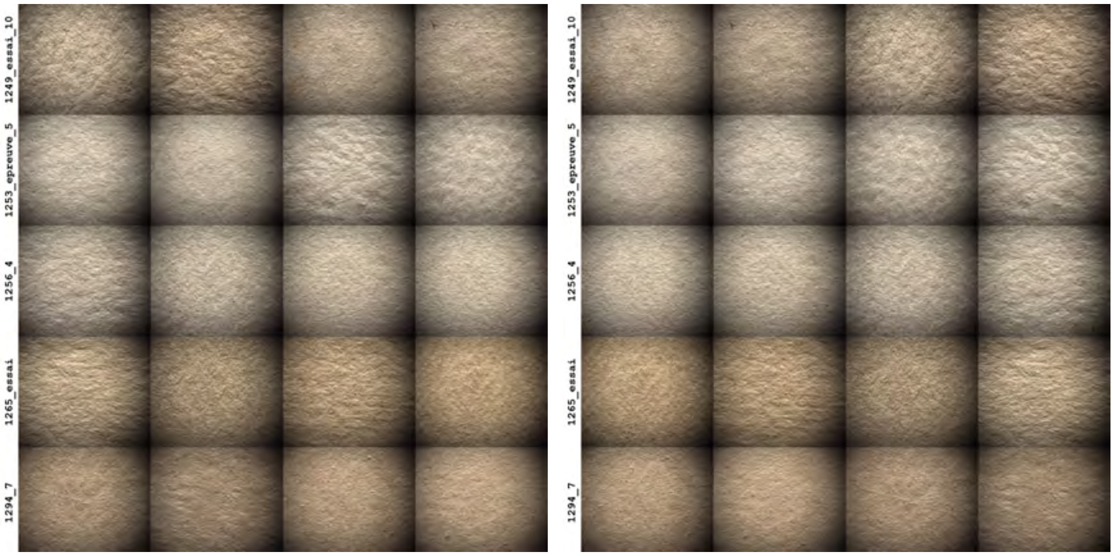
[3] P. Abry, A. Klein, W. Sethares, and C. Johnson, "Signal processing for art investigation: A shift to image feature mining (special issue editorial)," *IEEE Signal Processing Magazine*, vol. 32, pp. 14–16, Jul. 2015.



(a) Original permutation

(b) After applying a recto/verso flip to 2nd row

Fig. 3. Edition 1313 in both original and permuted order. Each row contains images corresponding to the four views of a single print.



(a) Original permutation

(b) After applying permutations $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{C}$, $\mathcal{A}$ to each of the five rows

Fig. 4. Representative papers from Editions 1249, 1253, 1256, 1265, and 1294, in both original and permuted order. Each row contains images corresponding to the four views of a single print from a different edition.

[4] 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.

[5] 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.

[6] C. R. Johnson, P. Messier, W. A. Sethares, A. G. Klein *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.

[7] 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.

[8] 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.

[9] 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.

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

[11] L. Lackey, A. Grootveld, and A. G. Klein, "Semi-supervised convolutional triplet neural networks for assessing paper texture similarity," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Nov. 2020.

[12] P. Abry, S. Roux, P. Messier, M. Ellis, and S. Jaffard, "Multiscale anisotropic analysis for assessment of similarity between papers in a large matisse print dataset," in *2020 54th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2020, pp. 137–141.

[13] 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.

[14] ASPECT Lab at WWU. (2021) Source code for paper texture classification and permutation of Matisse dataset. [Online]. Available: https://github.com/aspectlab/flippingmatisse