

Picasso, Matisse, or a Fake?

Automated Analysis of Drawings at the Stroke Level for Attribution and Authentication

Ahmed Elgammal
Artrendex LLC., NJ, USA
Department of Computer Science
Rutgers University, NJ, USA

Yan Kang
Artrendex LLC., NJ, USA

Milko Den Leeuw
Atelier for Restoration &
Research of Paintings (ARRS),
The Hague, The Netherlands

Abstract

This paper proposes a computational approach for analysis of strokes in line drawings by artists. We aim at developing an AI methodology that facilitates attribution of drawings of unknown authors in a way that is not easy to be deceived by forged art. The methodology used is based on quantifying the characteristics of individual strokes in drawings. We propose a novel algorithm for segmenting individual strokes. We propose an approach that combines different hand-crafted and learned features for the task of quantifying stroke characteristics. We experimented with a dataset of 300 digitized drawings with over 80 thousands strokes. The collection mainly consisted of drawings of Pablo Picasso, Henry Matisse, and Egon Schiele, besides a small number of representative works of other artists. The experiments shows that the proposed methodology can classify individual strokes with accuracy 70%-90%, and aggregate over drawings with accuracy above 80%, while being robust to be deceived by fakes.

Introduction

Attribution of art works is a very essential task for art experts. Traditionally, stylistic analysis by expert human eye has been a main way to judge the authenticity of artworks. This has been pioneered and made a methodology by Giovanni Morelli (1816-1891) who was a physician and art collector, in what is known as Morellian analysis. This connoisseurship methodology relies on finding consistent detailed “invariant” stylistic characteristics in the artist’s work that stay away from composition and subject matter. For example Morelli paid great attention to how certain body parts, such as ears and hands are depicted in paintings by different artists, not surprisingly given his medical background. This methodology relies mainly on the human eye and expert knowledge. The work of van Dantzig (van Dantzig 1973) that we follow in this paper belongs to this methodology.

In contrast, technical analysis focuses on analyzing the surface of the painting, the underpainting, and/or the canvas material. There is a wide spectrum of imaging (e.g. infrared spectroscopy and x-ray), chemical analysis (e.g. Chromatography), and radiometric (e.g. carbon dating) techniques that have been developed for this purpose. Mostly, this analysis aims to get insights on the composition of the materials and

pigments used in making the different layers of the work and how that relates to what materials, were available at the time of the original artist or what the artist typically used. These techniques are complementary and each of them has limitations to the scope of their applicability (Riederer 1986).

Analysis using computer vision and image processing techniques has been very sparsely and cautiously investigated in the domains of attribution and forgery detection (e.g. (Guo, Doermann, and Rosenfield 2000; Johnson et al. 2008; Polatkan et al. 2009; Li et al. 2012)). Image processing has been used as a tool in conjunction with non-visual spectrum imaging, such as analysis of x-ray imaging to determine canvas material and thread count (e.g. (Johnson et al. 2008; Liedtke, Johnson Jr, and Johnson 2012)).

The question we address in this paper is what role Artificial Intelligence can play in this domain given the spectrum of the other available technical analysis techniques, which might seem more conclusive. We argue that developing this technology would complement other technical analysis techniques for three reasons. First, computer vision can uniquely provide a quantifiable scientific way to approach the traditional stylistic analysis, even at the visual spectrum level. Second, it would provide alternative tools that lie out of the scope of applicability for the other techniques. For example, this can be very useful for detecting forgery of modern and contemporary art where the forger would have access to pigments and materials similar to what original artist had used). Third, computer vision has the potential to provide a cost-effective solution compared to the cost of other technical analysis methods. For example, in particular related to the topic of this paper, there are large volumes of drawings, prints, and sketches for sale and are relatively cheap (in the order of a few thousand dollars, or even few hundreds) compared to paintings. Performing sophisticated technical analysis in a laboratory would be more expensive than the price of the work itself. This prohibitive cost makes it attractive for forgers to extensively target this market.

Several papers have addressed art style classification, where style is an art movement (e.g. Impressionism), or the style of a particular artist (e.g. the style of Van Gogh) (Sablantig, Kammerer, and Zolda 1998; Fahad Shahbaz Khan 2010; Lombardi 2005; Arora and Elgammal 2012; Saleh et al. 2014). Such analysis does not target authentication. Such works use global features that mainly capture the composi-

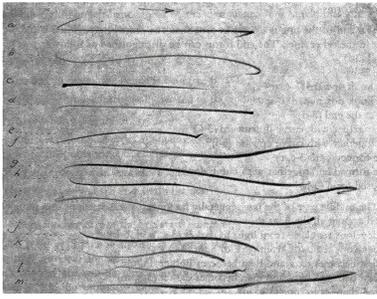


Figure 1: Illustration of van Dantzig methodology on simple strokes. Spontaneous strokes differ in their shape and tone at their beginning, middle and end. Figure from (van Dantzig 1973)

tion of the painting. In fact, such algorithm will classify a painting done on the style of Van Gogh, for example, as Van Gogh, since it is designed to do so.

Methodology The methodology used in this paper is based on quantifying the characteristics of individual strokes in drawings and comparing these characteristics to a large number of strokes by different artists using statistical inference and machine learning techniques. This process is inspired by the Pictology methodology developed by Maurits Michel van Dantzig (1903 - 1960). Van Dantzig suggested several characteristics to distinguish the strokes of an artist, and suggested that such characteristics capture the spontaneity of how original art is being created, in contrast to the inhibitory nature of imitated art (van Dantzig 1973).

Among the characteristics suggested by van Dantzig to distinguish the strokes of an artist are the shape, tone, and relative length of the beginning, middle and end of each stroke. The characteristics include also the length of the stroke relative to the depiction, direction, pressure, and several others. The list of characteristics suggested by van Dantzig is comprehensive and includes over one hundred aspects that are designed for inspection by the human eye. The main motivation is *to characterize spontaneous strokes characterizing a certain artist from inhibited strokes, which are copied from original strokes to imitate the artist style.*

In this paper we do not plan to implement the exact list of characteristics suggested by van Dantzig; instead we developed methods for quantification of strokes that are inspired by his methodology, trying to capture the same concepts in a way that is suitable to be quantified by the machine, is relevant to the digital domain, and facilitates statistical analysis of a large number of strokes by the machine.

We excluded using comparisons based on compositional and subject-matter-related patterns and elements. Most forged art works are based on copying certain compositional and subject-matter-related elements and patterns. Using such elements might obviously and mistakenly connect a test subject work to figures and composition in an artist known works. In contrast to subject matter and compositional elements, *the characteristics of individual strokes carry the artist's unintentional signature, which is hard to*

imitate or forge, even if the forger intends to do.

Contribution The main contribution of the paper is testing the hypothesis that artists can be identified based on individual strokes, which was never tested before scientifically. This is a novel problem and we provide a novel result, which shows that indeed the hypothesis is true. Beside this scientific contribution, technical novelty includes a novel stroke segmentation algorithm, designing and comparing different hand-crafted and learned deep neural network features, and adaptation of RNN for the task of quantifying stroke characteristics.

Detailed Methodology

Challenges The variability in drawing technique, paper type, size, digitization technology, spatial resolution, impose various challenges for quantify the characteristic of strokes in a way that is invariant to these variability. Here we highlight some these challenges and how we addressed them.

Drawings are made using different techniques, materials and tools, including, but not limited to drawings using pencil, pen and ink, brush and ink, crayon, charcoal, chalk, and graphite drawings. Different printing techniques also are used such as etching, lithograph, linocuts, wood cuts, and others. Each of these techniques results in different stroke characteristics. This suggests developing technique-specific models of strokes. However, typically each artist prefers certain techniques over others, which introduce unbalance in the data, which need to be addressed. Therefore, in this paper we are testing two hypotheses: technique specific vs. across technique comparisons, to test if we can capture invariant stroke characteristic for each artist that persists across techniques.

Drawings are executed on different types of papers, which, along with differences in digitization, imply variations in the tone and color of the background. This introduces a bias in the data. We want to make sure that we identify artists based on their strokes and not based on the color tone of the paper used. Different types of papers along with the type of ink used result in different diffusion of ink at the boundaries of the strokes which, combined with digitization effects, alter the shape of the boundary of the stroke.

Drawings are made on different-sized papers, and digitized using different resolutions. The size of the original drawing as well as the digitization resolution are necessary to quantify characteristics related to the width or length of strokes. Therefore, in this paper we quantify the characteristics of the strokes in a metric basis after converting all the measurements to the metric system.

Data collection A collection of 297 drawings were gathered from different sources. The drawings selected are restricted to line drawings, i.e, it excludes drawings that have heavy shading, hatching and water-colored strokes. The collection included drawings and prints by Picasso (130), Henry Matisse (77), Egon Schiele (36), Amedeo Modigliani (18), and a small representative works of other artists (36), ranging from 1910-1950AD. The collection included a variety of techniques including: pen and ink, pencil, crayon,

Table 1: Dataset collection: technique distribution

Technique	Pen/brush (ink)	Etching	Pencil	Drypoint	Lithograph	Crayon	Charcoal	Unknown	Total
Picasso	80	38	8	2	2	0	0	0	130
Matisse	45	10	5	2	14	1	0	0	77
Schiele	0	0	10	0	0	5	4	17	36
Modigliani	0	0	9	0	0	8	1	0	18
Others	20	0	0	0	9	4	1	2	36
Total	145	48	32	4	25	18	6	19	297
Strokes	36,533	19,645	9,300	914	6,180	4,648	666	2,204	80,090
Others: Georges Braque, Antoine Bourdelle, Massimo Campigli, Marc Chagall, Marcel Gimond, Alexej Jawlensky, Henri Laurens, Andre Marchand, Albert Marquet, Andr Masson, Andre Dunoyer Dr Segonzac, Louis Toughague									

and graphite drawings as well as etching and lithograph prints. Table 1 shows the number of drawings for each artist and technique. In the domain of drawing analysis it is very hard to obtain a dataset that is uniformly sampling artists and techniques. In all the validation and test experiments an equal number of strokes were sampled from each artist to eliminate data bias.

The collection included digitized works from books, downloaded digitized images from different sources, and screen captured images for cases where downloading was not permitted. The resolution of the collected images varies depending on the sources. The effective resolution varies from 10 to 173 pixel per cm depending on the actual drawing size and the digitized image resolution. Given this wide range of resolutions, the algorithms and features used were designed to be invariant to the digitization resolution.

Fake drawing dataset: In order to validate the robustness of the proposed approaches against being deceived by forged art, we commissioned five artists to make drawings similar to those of Picasso (24), Matisse (39) and Schiele (20) using the same techniques, totaling 83 drawings. None of these fake drawings was used in training the models. We only used them for testing.

Figure 2 shows examples of the fake dataset mixed up with real drawings. Because we do not expect the reader to be an expert in authentication in art, to be able to judge the quality of the fake drawings in isolation, we deliberately mixed up a collection of the fake drawings with real drawings in Figure 2. If the reader is interested to know which of these images are of fake or real drawings, please refer to the end of the paper.

Stroke Segmentation

A typical isolated stroke is a line or curve, with a starting point and endpoint. A stroke can have zero endpoints (closed curve) or 1 endpoint, which are special cases that do not need further segmentation. However, strokes typically intersect to form a network of tangled strokes that needs to be untangled. A network of strokes is characterized by having more than 2 endpoints. Since strokes are thin elongated structures; a skeleton representation would preserve their topological structure even in a network configuration (Lam, Lee, and Suen 1992). Therefore, the segmentation of strokes is done on such a skeleton representation.

There is a large classical literature in computer vision on detecting junctions on edge maps as a way to characterize object boundaries, infer about three-dimensional structure and form representations for recognition. Unlike classical literature which look at natural images, in our case detecting junctions and endpoints is fortunately relatively easy since they persists in a skeleton representation of the network of strokes. On the other hand, the challenge in our case is to use the information of such junctions and endpoints to segment individual strokes.

In our case, junctions play crucial role in identifying the intersections between strokes. There are two basic ways strokes intersect: occluder-occluded configuration to form a T-junction or two strokes crossing each other to form an X-junction. A T-junction is a continuation point of the occluding stroke and an endpoint for the occluded stroke. We need to preserve the continuation of the occluding stroke at the T-junction.

The stroke segmentation algorithm takes a network of strokes and identifies one occluding stroke at a time and remove it from the network of strokes to form a residual network(s) that is recursively segmented. This is achieved by constructing a fully connected graph whose vertices are the endpoints in the network and edges are weighted by the cost of reaching between each two endpoints. The cost between two endpoints reflects the bending energy required at the junctions.

Let the endpoints in a network of strokes denoted by e_1, \dots, e_m and let the junction locations denoted by j_1, \dots, j_n . The cost of the path between any two end points e_i to e_j is cumulative curvature along the skeleton path between them, where the curvature is only counted close to junctions. The rational is that it does not matter how much bending a stroke would take as long as it is not at junctions. Let $\gamma(t) : [0 : 1] \rightarrow R^2$ be the parametric representation of the curve connecting e_i and e_j . The cost is defined as

$$c(e_i, e_j) = \int_0^1 \kappa(t) \cdot \phi(\gamma(t)) dt$$

where $\kappa(\cdot)$ is the curvature and $\phi(\cdot)$ is a junction potential function, which is a function of the proximity to junction locations defined as

$$\phi(x) = \frac{1}{n} \sum_{i=1}^n e^{(x-j_i)^2/\sigma}$$



Figure 2: Examples of images of the fake dataset mixed up with real images of drawings by Matisse, Picasso, and Schiele. See the key at the end of the document to tell which are real and which are fake!

After the graph construction, the minimum cost edge represents a path between two endpoints with minimum bending at the junctions, which corresponding to an occluding stroke. In case of a tie, the path with the longest length is chosen. The optimal stroke is removed from the skeleton representation and from the graph. This involves reconnecting the skeleton at X-junctions (to allow the detection of the crossing strokes) and new endpoints have to be added at T-junctions (to allow the detection of occluded strokes). Removing a stroke from the graph involves removing all edges corresponding to paths that go through the removed stroke. This results in breaking the graph to one or more residual subgraphs, which are processed recursively.

Stroke Analysis Methodology

Quantifying Stroke Characteristics

This section explains the process of quantifying the characteristics of individual strokes and the extracted features used to represent each stroke. The goal is to construct a joint feature space that captures the correlation between the shape of the stroke, its thickness variation, tone variation, local curvature variation. For this purpose we studied two different types of features and their combination: 1) Hand-crafted features capturing the shape of each stroke and its boundary statistics, 2) Learned-representation features capturing the tone variation as well as local shape characteristics. The next two subsection describe each of these features.

Hand-crafted Features In our study, each stroke is represented by its skeleton, its boundary, and the rib length around the skeleton. The following descriptors are extracted to quantify the characteristics of each stroke. All the descriptors are designed to be invariant to translation, rotation, scaling, and change in digitization resolution.

Shape of the boundary: The shape of the stroke boundary is quantified by Fourier descriptors, which provide shape features that are proven to be invariant to translation, scaling, rotation, sampling, and contour starting points (Burger and Burge 2016). We used 40 amplitude coefficients (first 20 harmonics in each direction) to represent the shape of the boundary of the stroke.

Reconstruction error profile: The mean reconstruction error, as a function of the number of harmonics used to approximate the shape of the strokes, is used as a descriptor of the smoothness of the contour and the negative space associated with the stroke. In particular, we compute the mean reconstruction error at each step while incrementally adding more harmonics to approximate the shape of the stroke. The reconstruction error profile is normalized by dividing by the stroke mean width in pixels to obtain a descriptor invariant to digitization resolution.

Contour Curvature descriptor: To quantify the curvature of the stroke contours, we use the first and second derivatives of the angular contour representation. The distributions of these derivatives are represented by their histograms.

Stroke thickness profile: To quantify the thickness of the stroke, we compute the mean and standard deviation of the rib length around the skeleton of the stroke, as well as a histogram of the rib length. All rib length measurements are

mapped to mm units to avoid variations in digitization resolution.

Stroke Length: The length of the stroke is quantified as the ratio between the stroke skeleton length to the canvas diagonal length. This measure is invariant to digitization resolution.

Deep Learned Features using RNNs *GRU Classification with Truncated Back Propagation Through Time:*

Unlike traditional feed-forward neural networks specialized at fixed size input, e.g. images, recurrent neural network (RNN) could handle variable length sequence and either fixed length output or variable length outputs. RNN sequentially takes input from the input sequence and update its hidden state. Recently, it has been widely shown that more complicated RNN model such as Long Short-Term Memory (LSTM, (Hochreiter and Schmidhuber 1997)) or Gated Recurrent Unit (GRU, (Chung et al. 2014)) would eliminate the problem of vanishing gradient (Bengio, Simard, and Frasconi 1994), (Hochreiter et al. 2000). LSTM and GRU introduce some gating units that can automatically determine how much the information flow could be used in each time step, by which the vanishing gradient could be avoided.

We adapt GRUs for the task of quantifying individual stroke tone and local shape features. Given a stroke, a sequence of patches of fixed size are collected along the skeleton of the stroke and fed to a GRU model as inputs. We tested both fixed size patches or adaptive size patches where the radius of the patch is a function of the average stroke width in the drawing. In both cases the input patches are scaled to 11x11 input matrices. To achieve invariant to the direction of the stroke, each stroke is sampled in both directions as two separate data sequences (at classification, both a stroke and its reverse either appear in training or testing splits). We normalized the grey scale into range (-1, 1), and flattened the 11 × 11 image into a 121-dimension vector. The activation function we used in experiments is *tanh* function. Parameters are initialized from normal distribution with mean = 0, standard deviation = 1. After comparing several optimizer functions, we found that the RMSProp optimizer with learning rate 0.001 outperforms others. To avoid gradient vanishing, we calculated the gradient by the truncated Back Propagation Through Time. Each sequence is unrolled into a fixed size steps (30 in the experiments) at each time to calculate the gradient and to update the network's parameters. The label of original sequence is assigned to each unrolling. Between each unrolling, the hidden state is passed on to carry former time steps information. Within each unrolling, only the last time step hidden state is used in the final linear transformation and Softmax function to get the predicted score of each class. We used a cross entropy loss.

Stroke Classification

For the case of hand-crafted features, strokes are classified using a support vector machine (SVM) classifier (Cortes and Vapnik 1995). We evaluated SVM using Radial basis kernels as well as polynomial kernels. The classifier produces posterior distribution over the classes. For the case of learned GRU features, the classification of strokes is directly given



Figure 3: Examples of Segmentation Results — best seen in color.

by the trained networks. SVM was used to combine hand-crafted features with the learned features in one classification framework. In such case, the activation of the hidden units were used as features, and combined to the hand-crafted features.

Drawing classification:

A given drawing is classified by aggregating the outcomes of the classification of its strokes. We used four different strategies for aggregating the stroke classification results, as described below. 1) *Majority Voting*: In this strategy each stroke votes for one class. All strokes have equal votes regardless of the certainty of the output of the stroke classifier. 2) *Posterior aggregate*: In this strategy each stroke votes with a weight equal to its posterior class probability. This results in reducing the effect of strokes that are not classified with high certainty by the stroke classifier. 3) *k-certain voting*: In this strategy, only the strokes with class posterior greater than a threshold k are allowed to vote. This eliminates effect of uncertain strokes. 4) *certainty weighted voting*: In this strategy each stroke vote is weighted using a gamma function based on the certainty of the stroke classifier in classifying it.

Validation Experiments

This section describes the experiments conducted to test and validate the performance of the proposed stroke segmentation, stroke classification, and the drawing classification approaches on the collected dataset. In particular, the experiments are designed to test the ability of the algorithms to determine the attribution of a given art work and test its robustness to forged art. A baseline comparing to a holistic image-level approach is available.

Segmentation Validation

Validating the segmentation algorithm is quite challenging since there is no available ground truth segmentation and because of the difficulty of collecting such annotation. It is quite a tedious process for a human to trace individual strokes to

provide segmentation of them, specially such task requires certain level of expertise. To validate the segmentation algorithm, we collected 14 drawings with medium difficulty (in terms of number of strokes) from the collection and showed the segmentation results to two artists and asked them independently to locate errors in the segmentations. Figure 4 shows an example of a drawing with its two annotations of the results. A closer look highlights that annotators make several mistakes (false positive, and false negatives). The overall error per annotator is computed as: Error rate = total marked errors at junctions / total number of strokes; where the total is aggregated over all evaluated images. The average error rate over the two annotators is 12.94%, counting all labeled errors by annotators (see SM for details). The annotation shows large deviations between the two annotators, with mean deviation 24.93 and standard deviation 12%. This highlights the challenge in validating the segmentation results by human annotation. However, most of the marked errors are at small detailed strokes that are hard to segment, even by the human eye, and does not contribute much to the classification of strokes since small strokes are filtered out anyway.

Stroke Classification

Evaluation Methodology: In all experiments the image datasets were split into five 80/20% folds to perform five-fold cross validation. Since strokes from the same drawings might share similar characteristics, we did this splits at the image level and not at the stroke level. For each fold, after splitting the images to train and test sets, equal number of strokes were sampled for each artist class for training and testing to avoid the bias in the data, which is significant in our case. We evaluated different classification settings including pair-wise classification, and one-vs-all classification, and multi-class classification. Extensive ablation studies are also performed to evaluate the different features and their effects, as well as to choose the optimal settings.

Stroke Classification Validation - Technique Specific - Pairwise: For testing technique-specific classifiers, we

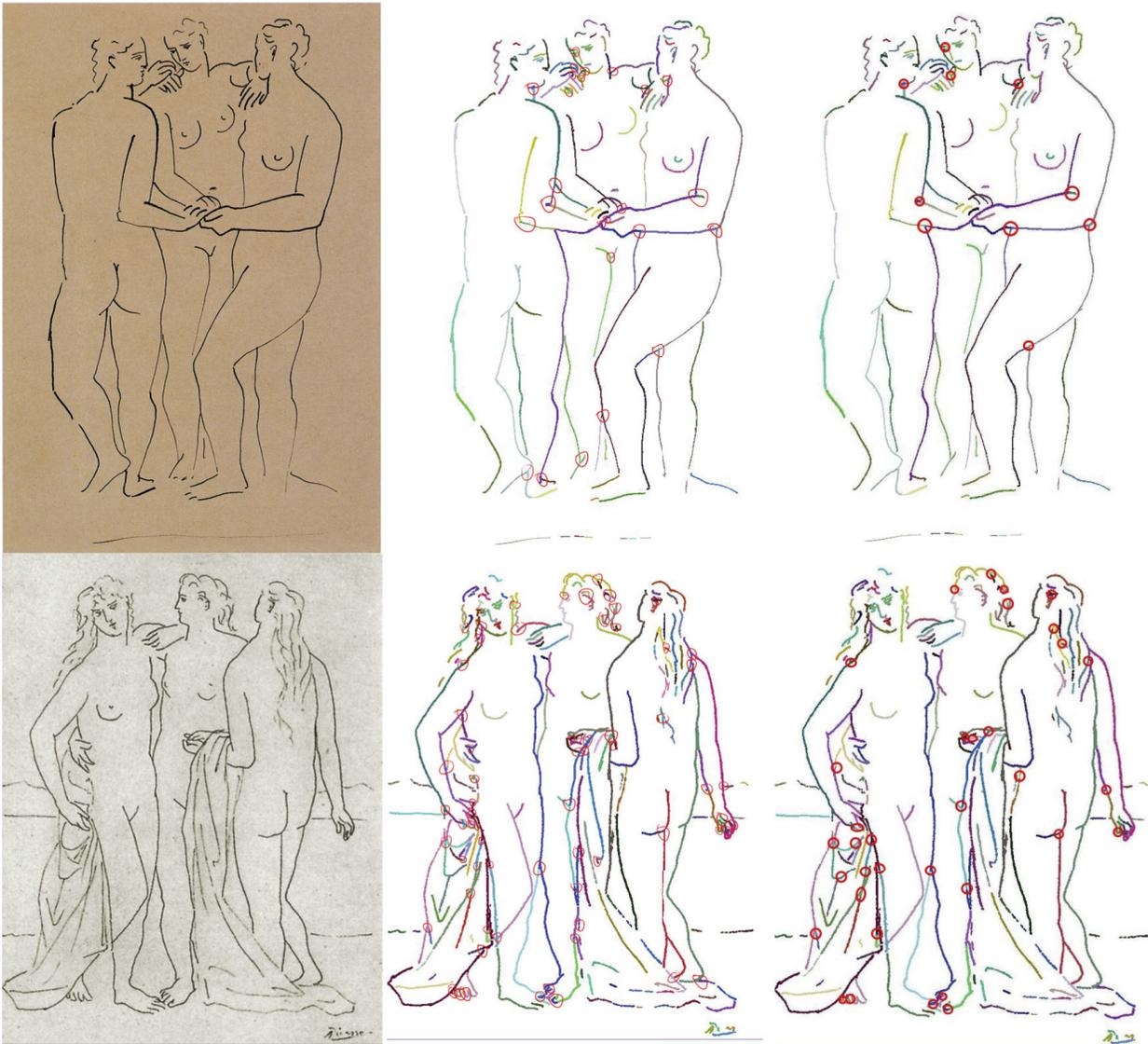


Figure 4: Example of two drawings by Picasso, stroke segmentation results, and segmentation errors marked by two artists.

trained pairwise classifiers to discriminate between Picasso and Matisse drawings made using either pen/ink or etching. We chose these two techniques and these two artists since they have the largest representation in our collection. Table 2-top shows the stroke classification results. The experiments is done using five-fold cross validation and the mean and standard deviations are reported. The table shows a comparison between the different types of proposed features.

Stroke Classification Validation - One-vs-all: In this experiment a one-vs-all classification settings is used to build classifiers for Picasso-vs-Non-Picasso, Matisse-vs-Non-Matisse, Schiele-vs-Non-Schiele. These three artists are chosen since they have enough data for training and testing the classifiers in a five-fold split setting. The classifiers are then evaluated on the fake dataset.

We evaluated the performance of two settings: 1) across-

techniques: we evaluated the performance of the stroke classifiers on all techniques combined to evaluate whether the classifier can capture an invariant for the artist regardless of the technique used. 2) Technique-specific: in this setting each classifier is trained and tested using strokes from the same drawing technique. Given the data collection, we tested a) Picasso-vs-Non-Picasso classifier using ink/pen, b) Matisse-vs-Non-Matisse classifier using ink/pen, c) Schiele-vs-Non-Schiele using pencil.

Table 3-top shows the mean and standard deviations of the five folds for the hand-crafter features, the GRU features and the combination. Both types of features have very good stroke classification performance. GRU has better performance over the three artists tested. Combining the features further improved the results and reduced the cross-fold variances, which indicate that both types of features are comple-

Table 2: Validation of Stroke Classifier

Technique specific: Ink Drawing (Pen/Brush) (Picasso vs Matisse)		
Approach	Train	Test
Hand-Crafted - SVM-RBF	87.99% (0.39%)	79.16% (0.26%)
Hand-Crafted - SVM-POLY	79.88% (0.14%)	77.17% (0.58%)
GRU	84.92% (1.89%)	65.86 (13.58 %)
Etching Prints (Picasso vs Matisse)		
Approach	Train	Test
Hand-Crafted - SVM-RBF	94.53% (0.22%)	84.18% (0.85 %)
Hand-Crafted - SVM-POLY	94.27% (0.21%)	93.09% (0.88%)
GRU	83.74% (4.60 %)	75.08% (8.11%)
Multiclass - all techniques		
5 Classes: Picasso, Matisse, Schiele, Modigliani, Others		
Approach	Train	Test
Hand-Crafted - SVM	55.01 % (1.41%)	48.97% (5.82%)
GRU	87.72% (2.43%)	74.65% (3.41%)

mentary to each other as we hypothesized.

Comparing the performance of stroke classifiers on both the technique-specific and across-technique settings, we notice that in both cases the classifiers performed well. The GRU performed better in the across-technique settings than in the technique-specific setting, which can be justified by the lack of data in the later case.

Multi-class Stroke Classification Experiment See Table 2-bottom.

Drawing Classification and Detection of Fakes

Drawing Classification Validation: The trained one-vs-all stroke classifiers were tested for drawing classification using the four aforementioned strategies for aggregating the results from the stroke level to the drawing level. Given that the stroke classifiers are trained on a five-fold cross-validation setting, the drawing classification followed that strategy, i.e. in each fold, each drawing in the test split is classified using the classifier trained on the 80% of the images in the training split, hence there is no standard deviation to report. Table 3-bottom shows the results for the across-technique setting and the technique-specific setting.

Evaluation on Fake Drawings: The trained stroke classifiers were also tested on the fake drawing set to evaluate whether the classifiers are really capturing artists’ stroke characteristics and invariants or just statistics that can be easily deceived by forged versions. We used the Picasso-vs-all classifiers to test the fakes that are made to imitate Picasso drawings (We denote them as Picasso fakes). A similar setting is used for Matisse fakes and Schiele fakes. Since the stroke classifiers are trained on a five-fold setting, we have five different classifiers trained per artist, one for each fold. Each test stroke is classified using the five classifiers and the majority vote is computed. The different aggregation methods are used to achieve a final classification for each drawing. Since one-vs-all setting is adapted, classifying a fake Picasso as others in a Picasso-vs-all setting is considered a correct classification, while classifying fake Picasso as Picasso is considered a wrong prediction. The bottom parts of Table 3 shows the results for the fake dataset for the across-technique and technique-specific settings. The table

shows that the trained one-vs-all stroke classifiers for all the three artists, are robustly rejecting fake drawing with accuracy reaching 100% in the across-technique case. A notable difference here is that the GRU failed to detect the fake drawings, in particular for the Picasso-vs-all, while the hand-crafted features detected all the fakes. Similar case happens for Schiele-vs-all as well. We hypothesize that this is because of the limited training data in the technique-specific case, which did not allow the GRU to learn an invariant model that generalizes well as in the across-technique case. In contrast the hand-crafted models did not suffer from this limitation. Overall, the hand-crafted features are outperforming in detecting the fakes.

Conclusion

In this paper we proposed an automated method for quantifying the characteristics of artist strokes in drawings. The approach is inspired by the Pictology methodology proposed by van Dantzig. The approach segments the drawing into individual strokes using a novel segmentation algorithm. The characteristics of each stroke is captured using global and local shape features as well as a deep neural network that captures the local shape and tone variations of each stroke. We compared different types of features and showed results at the stroke classification and drawing classification levels.

The main result of this paper is that it shows that we can discriminate between artists at the stroke-level with high accuracy, even using images of drawing of typical off-the-web or scanned books resolutions. We also tested the methodology using a collected data set of fake drawings and the results show that the proposed method is robust to such imitated drawings, which highlights that the method can indeed capture artists’ invariant characteristics that is hard to imitate.

References

- Arora, R. S., and Elgammal, A. M. 2012. Towards automated classification of fine-art painting style: A comparative study. In *ICPR*.
- Bengio, Y.; Simard, P.; and Frasconi, P. 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Nets* 157–166.
- Burger, W., and Burge, M. J. 2016. *Fourier Shape Descriptors*. London: Springer London. 665–711.
- Chung, J.; Gulcehre, C.; Cho, K.; and Bengio, Y. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *NIPS’2014 Deep Learning workshop*.
- Cortes, C., and Vapnik, V. 1995. Support-vector networks. *Machine Learning* 20(3):273–297.
- Fahad Shahbaz Khan, Joost van de Weijer, M. V. 2010. Who painted this painting?
- Guo, J. K.; Doermann, D.; and Rosenfield, A. 2000. Off-line skilled forgery detection using stroke and sub-stroke properties. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, 355–358. IEEE.
- Hochreiter, S., and Schmidhuber, J. 1997. Long short-term memory. *Neural Computation* 9(8):1735–1780.

Table 3: Validation of Stroke-level and Drawing-level Classifiers

Validation of Stroke Classifiers - One-vs-All									
Across-Techniques - Mean (std) of five folds									
Classifier	technique	Hand-crafted + SVM			GRU			Combined	
		train	test		train	test		train	test
Picasso vs. all	All	72.59% (1.19 %)	67.26% (8.37 %)		81.92% (2.59 %)	75.09% (5.09%)		86.05% (1.08 %)	78.54% (4.36 %)
Matisse vs. all	All	65.83% (1.72 %)	60.61% (8.71%)		81.01% (3.41 %)	72.68% (5.58 %)		87.92% (1.73%)	77.08% (4.33%)
Schiele vs. all	All	84.76% (0.91%)	81.49% (3.30 %)		85.55% (1.74 %)	78.54% (8.77%)		91.85% (0.87 %)	86.20 % (3.78%)
Technique-specific - Mean (std) of five folds									
Classifier	Technique	Hand-crafted + SVM			GRU			Combined	
		train	test		train	test		train	test
Picasso vs. all	Pen/Ink	73.20% (2.21%)	68.93% (7.04%)		84.08% (2.20%)	72.24% (1.87 %)		88.40% (1.19%)	75.92% (4.22%)
Matisse vs. all	Pen/Ink	73.35% (1.99 %)	70.08% (7.94%)		86.88% (1.98 %)	75.03% (6.47%)		91.56% (1.03%)	79.10% (6.65 %)
Schiele vs. all	Pencil	82.58% (2.78%)	75.39% (20.64 %)		94.33% (3.52 %)	69.60% (20.62 %)		91.30% (4.57 %)	72.93% (19.67 %)

Validation of Drawing Classifiers - - One-vs-All									
Across-Techniques									
Aggregation	Picasso-vs-All			Matisse-vs-All			Schiele-vs-All		
	Hand-crafted	GRU	Combined	Hand-crafted	GRU	Combined	Hand-crafted	GRU	Combined
Majority	66.67%	76.77%	82.49%	54.88%	81.14%	80.47%	74.41%	82.49%	81.82%
Posterior	67.68%	77.44%	81.48%	56.90%	81.48%	79.12%	74.75%	83.50%	82.49%
85%-certain	73.06%	79.80%	82.83%	38.05%	80.47%	78.79%	75.42%	83.50%	83.84%
Certainty-weighted	67.34%	79.80%	82.83%	58.25%	80.81%	80.47%	75.42%	85.19%	83.16%
Detection of Fakes*	100%	87.50 %	100%	76.92 %	100%	100%	100%	100%	100%
Technique-specific									
Aggregation	Picasso-vs-All			Matisse-vs-All			Schiele-vs-All		
	Hand-crafted	GRU	Combined	Hand-crafted	GRU	Combined	Hand-crafted	GRU	Combined
Majority	72.41%	82.76%	81.38%	65.52%	78.62%	82.76%	81.25%	78.12%	81.25%
Posterior	72.41%	82.76%	81.38%	66.21%	79.31%	80.69%	84.38%	78.12%	81.25%
85%-certain	72.41%	82.76%	82.76%	69.66%	76.55%	80.69%	84.38%	78.12%	81.25%
Certainty-weighted	71.72%	82.76%	82.07%	69.66%	77.93%	80.00%	87.50%	78.12%	81.25%
Detection of Fakes*	100.00%	12.50%	20.83%	97.44%	100.00%	100.00%	100.00%	45.00%	60.00%

* All aggregation methods resulted in almost the same results for the fake detection. Due to limited space we only report the Certainty-weighting results.

Hochreiter, S.; Bengio, Y.; Frasconi, P.; and Schmidhuber, J. 2000. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. *Field Guide to Dynamical Recurrent Networks*.

Johnson, C. R.; Hendriks, E.; Berezhnoy, I. J.; Brevdo, E.; Hughes, S. M.; Daubechies, I.; Li, J.; Postma, E.; and Wang, J. Z. 2008. Image processing for artist identification. *IEEE Signal Processing Magazine* 25(4).

Lam, L.; Lee, S.-W.; and Suen, C. Y. 1992. Thinning methodologies-a comprehensive survey. *IEEE Transactions on pattern analysis and machine intelligence* 14(9):869–885.

Li, J.; Yao, L.; Hendriks, E.; and Wang, J. Z. 2012. Rhythmic brushstrokes distinguish van gogh from his contemporaries: findings via automated brushstroke extraction. *IEEE transactions on pattern analysis and machine intelligence* 34(6):1159–1176.

Liedtke, W.; Johnson Jr, C. R.; and Johnson, D. H. 2012. Canvas matches in vermeer: A case study in the computer analysis of fabric supports. *Metropolitan Museum Journal* 47(1):101–108.

Lombardi, T. E. 2005. The classification of style in fine-art painting. *ETD Collection for Pace University. Paper AAI3189084*.

Polatkan, G.; Jafarpour, S.; Brasoveanu, A.; Hughes, S.; and Daubechies, I. 2009. Detection of forgery in paintings using supervised learning. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, 2921–2924. IEEE.

Riederer, J. 1986. The detection of art forgeries with scien-

tific methods. In *Forensic science progress*. Springer. 153–168.

Sablatnig, R.; Kammerer, P.; and Zolda, E. 1998. Hierarchical classification of paintings using face- and brush stroke models. In *ICPR*.

Saleh, B.; Abe, K.; Arora, R. S.; and Elgammal, A. 2014. Toward automated discovery of artistic influence. *Multimedia Tools and Applications* 1–27.

van Dantzig, M. M. 1973. *Pictology: An Analytical Method for Attribution and Evaluation of Pictures*. E. J. Brill, Leiden, Netherlands.

Figure 2 key:
 Fake, Fake, Matisse
 Matisse, Fake, Fake, Matisse
 Fake, Matisse, Picasso, Fake
 Fake, Picasso, Picasso, Fake
 Schiele, Fake, Fake, Schiele, Schiele, Fake