Computational identification of significant actors in paintings through symbols and attributes

David G. Stork,[a] Anthony Bourached,[b] George H. Cann,[c] and Ryan-Rhys Griffiths[d]

[a]Portola Valley, CA 94028 USA
[b]Oxia Palus
[c] Department of Space and Climate Physics, University College London, London, UK
[d]Department of Physics, Cambridge University, Cambridge, UK

ABSTRACT

The automatic analysis of fine art paintings presents a number of novel technical challenges to artificial intelligence, computer vision, machine learning, and knowledge representation quite distinct from those arising in the analysis of traditional photographs. The most important difference is that many realist paintings depict stories or episodes in order to convey a lesson, moral, or *meaning*. One early step in automatic interpretation and extraction of meaning in artworks is the identifications of figures ("actors"). In Christian art, specifically, one must identify the actors in order to identify the Biblical episode or story depicted, an important step in "understanding" the artwork. We designed an automatic system based on deep convolutional neural networks and simple knowledge database to identify saints throughout six centuries of Christian art based in large part upon saints' symbols or *attributes*. Our work represents initial steps in the broad task of automatic semantic interpretation of messages and meaning in fine art.

Keywords: computational art analysis, artificial intelligence, computer-assisted connoisseurship, religious symbols and attributes, deep neural networks, semantic image analysis, visual semiotics

1. INTRODUCTION AND BACKGROUND

The problem of even basic automatic interpretation of fine art paintings and drawings is extremely challenging and is quite different from semantic image analysis of traditional photographs. In most current research on automatic segmentation, object recognition, image captioning, question answering, and analysis of photographs, the output describes the state of affairs in a physical scene. Many paintings, drawings and murals, particularly in the Western canon, were created to tell stories, convey lessons, or impute morals that have no counterpart in the vast research on analysis of natural photographs. Much narrative art—and specifically most Western religious art—depicts episodes or stories that have a crafted or selected *meaning*. Such meanings are expressed in natural language, as for instance "Christ died for your sins," "You should be willing to sacrifice your own son if your god demands it," "Extend charity to those less fortunate," "The love of money is the root of all evil," and innumerable others. Full semantic analysis of such artwork must eventually extract or infer such meanings.

Artists often employ conventions, deliberate physical inconsistencies, highly stylized renderings, and more in order to help convey such messages and meanings. Most artists, especially in the last millennium, use content, composition, medium, color, style, visual metaphors, and conventions such as symbols, "invisible" entities, and other techniques in the service of these goals. Many of these properties of art have no counterpart in traditional natural photographs or videos and for this reason the techniques for the analysis of common photographs are of only modest value in automatic approaches to interpreting fine art.

The understanding of meaning in much representational Western art, such as religious art, builds upon the identification of depicted figures or "actors." For instance, the meaning in Vincenzo Catena's *Christ giving the keys to Saint Peter* is lost on viewers who cannot recognize the central figures, that the keys symbolize the requirements for entry into heaven, and that the three women in the background are allegories of Faith, Hope, and Charity. Traditional automatic facial recognition is nearly useless in this task, as there is no "ground

---

Send correspondence to the first author at artanalyst@gmail.com.

truth" exemplar images of important figures. Moreover, the stylistic variations among painters are far too great for current image recognition systems trained with natural photographs. No current face recognition system is accurate for faces as diverse as those that appear in the works of Duccio, Leonardo, El Greco, Karel Apel, Jean Dubuffet, Willem de Kooning, among many others. Art analysis differs methodologically from that for traditional photographs as well in that the total number of artworks available for training automatic systems, while large, is but a small fraction of the number of photographs available online and in curated databases that are vital to the success of state-of-the-art performance based on training deep neural networks.[1] This disparity in corpora sizes is growing every day.

Central to the understanding of religious art such as Christian art in the Western canon are *symbols*, for instance a cross, lamb, keys, crown of thorns, ox and ass, and so on. Furthermore, the non-physical conditions such as the upward floating of Christ in Rogier van der Weyden's *Isenheim altarpiece*, the glowing radiant Christ child in Gerard David's *Birth of Christ*, the halos appear in innumerable artworks, all conveying important and well-understood meanings. Conventions such as the appearance of crepuscular rays ("Rays of Buddha") and golden heavenly rays have associations and thus convey meaning. For all its aesthetic value, Rembrandt's *Supper at Emmaus* would appear as a common dinner in a simple inn were it not for the radiant glow of the central figure, indicating his identity as Christ, and thus marking the scene as one of the central stories in the Bible. One simply cannot understand such artworks and the artists' intentions without recognizing and reasoning about these conventions and their associations.[2,3] Any automatic system for the interpretation of art must recognize and then make inferences using these unique artistic aspects.

One of the requisite early steps in interpreting such paintings, and thus automatic interpretation, is identifying the figures or "actors," such as saints in Bible stories. This is the task we address: the automatic identification of major figures or "actors," such as saints, in religious two-dimensional art, specifically paintings. Major figures such as saints in Western religious art are frequently marked by the presence of a sign or *attribute*.[4] The general categories of *signs*—so-called *signifiers*—and their corresponding objects or concepts—so-called *signifieds*—is the central concern of the branch of philosophy known as *semiotics*.[5,6] Thus Christ's main attributes are the cross and the lamb (*Agnus Dei*), Saint Peter's are keys, Saint Mark's is the winged lion, God's is a dove, and so on. In religious iconography, an *attribute* is an object or symbol associated with a figure such as a saint, most often deriving from some episode or story from the Bible. For instance, Saint Peter's attribute, *keys*, stems from a passage in *Matthew* 16:19:

> "I will give to thee the keys of the kingdom of heaven. And whatsoever thou shalt bind upon earth, it shall be bound also in heaven: and whatsoever thou shalt loose upon earth, it shall be loosed also in heaven."

Some religious actors have several attributes (Saint Peter: keys, boat, and fish), and several actors share the same attribute (Christ and John the Baptist: cross), though which of multiple attributes is represented depends upon the settings, context, or episode being depicted. Such symbols and attributes were widely used in religious paintings, altarpieces, stained-glass windows, prayer books, and so on, to help teach and reinforce Biblical stories, particularly to illiterate parishioners.[4]

Figure 1 shows five paintings of Saint Peter from different art historical periods, all showing his attribute, keys. All viewers familiar with Christian iconography recognize that these paintings depict the same figure, despite significant differences in size, style, lighting, pose, and scene, all because of the presence of the attribute, itself portrayed amidst numerous size and stylistic variations. While the figure's identity can also be indicated by his visage and beard, costume (papal vestments and pallium), and in some cases setting, such visual properties can differ widely for a given saint and be quite similar for different saints. In short, a symbol or attribute is often the most a reliable visual marker of many important saints in Christian art.

We note in passing that most mythical and religious art of nearly all religions and sects employs such symbols functioning as attributes do. In Greek and Roman mythology, for instance, Cupid is symbolized by a small bow and arrow, Zeus by a lightning bolt, Athena by an owl, Hera by a peacock, Poseidon by a trident, Bacchus by a bunch of grapes, and so forth.[7] Most of these associations have roots in Ovid's *Metamorphoses*. Likewise in Hindu iconography Brahma is represented by a lotus, Saraswati by a veena (stringed musical instrument),

Figure 1. a) Unknown artist's *Stefani Triptych, Saint Peter enthroned* (c. 1330), b) unknown artist's *Saint Peter*, Convent of Saint Mary, Zadar (12th century), c) Girolamo Dai Libri's *Saint Peter with keys* (1533), d) Perugino's *Delivery of the keys to Saint Peter*, Vatican (1482), and e) Peter Paul Rubens' *Saint Peter* (c. 1611). In works such as these, Saint Peter is recognized by the presence of his attribute, keys.

Parvati by a lion, and so on. Most of these symbols have roots in ancient Hindu texts such as the *Bhagavad Gita*.

We demonstrate an automatic method for identifying actors—as a step toward inferring the source story and hence meaning—in Western Biblical art. We present background for this goal and approach in Sect. 2. Then, in Sect. 3 we discuss the collection of religious art images for analysis and in Sect. 4 an overview of our method for classifying symbols and attributes within paintings. Next, in Sect. 5 we mention briefly our database of associations between attributes and saints. In Sect. 6 we discuss the training and testing of our deep neural networks for recognizing and localizing attributes and segmenting figures. We quantify our results in Sect. 7 and interpret the sources of inevitable errors. In Sect. 8 we summarize our conclusions and discuss next steps in extraction of meaning from artwork and related problems in high-level image analysis of art.

To the best of our knowledge, the work presented here is the first explicit work on automatic recognition of significant actors in art by any means, specifically through the use of symbols or attributes.

2. TECHNICAL GOAL AND APPROACH

In broad overview, our goal is to develop an automated system of image analysis that can, given a Christian painting in the Western canon from the past six centuries, identify saints (or major "actors"). Our system relies on an ATTRIBUTE RECOGNITION MODULE for symbols and attributes, a SEMANTIC SEGMENTATION MODULE, to localize candidate figures (actors), and an ATTRIBUTE SAINT ASSOCIATION DATABASE, a database of attributes (signifiers) and association saints (signifieds).

Our approach includes the following steps:

- train a deep neural network ATTRIBUTE RECOGNITION MODULE to recognize and localize attributes (religious symbols associated with saints) from a small set of common such attributes, as listed in Table 1.

Given a painting to be analyzed:

- use a SEMANTIC SEGMENTATION MODULE (deep neural network) to segment the painting image, specifically to identify human actors

- use the trained ATTRIBUTE RECOGNITION MODULE (deep neural network) to identify attributes and their locations within the artwork

- find the nearest segmented human figure to each identified attribute

- use the ATTRIBUTE SAINT ASSOCIATION DATABASE to identify the actor based on its closest semiotic attribute

3. IMAGE COLLECTION AND DATA PREPARATION

We scraped images of fine art Christian paintings from the 13th through 19th centuries in major museums, hand picked using search terms such as saints' names, the attributes, "painting," "mural," "fresco," "altarpiece," "Medieval" "Renaissance," "Baroque," and so forth. We confirmed by eye that indeed each work depicted at least one saint and at least one attribute. We hand classified attributes but not actors (saints). We scaled all images to the same total number of pixels, to match the number of neurons in the input and output layers of our network (see Sects. 4–6).

4. ATTRIBUTE RECOGNITION MODULE

We trained a well-studied the convolutional architecture,[8] which stacks a feature pyramid network on top of a deep residual network[?] for region of interest extraction and binary pixel classification and segmentation. We used the pre-trained weights trained by on the open-source common objects in context dataset.[?,8] For instance, we trained the network with images of the "keys to heaven" depicted in the art in question.

5. ATTRIBUTE SAINT ASSOCIATION DATABASE

We created our attribute database by hand, using scholarly sources of Christian iconography.[4] Table 1 shows the several saints and attributes used in our study as the ATTRIBUTE SAINT ASSOCIATION DATABASE. This list is of course not complete, either in the number of saints or the full complement of attributes for some saints.

Saint	Attribute
Christ	cross
Matthew	angel
Mark	winged lion
Luke	bull
Simon	boat
Thomas	ax
Catherine	wheel
Daniel	lion
George	dragon
John	eagle

Table 1. Several saints frequently depicted in Western Christian art and their leading attributes, stored as associations in our ATTRIBUTE ASSOCIATION DATABASE. Several of the most-important saints have multiple attributes, for instance, Saint Peter's attributes include keys of heaven, fish, boat, rooster, pallium and papal vestments, as discussed in Sect. 8.

6. TRAINING AND TESTING

We used well-established protocols of leave-one-out training to ensure that any test image was not used during training.[9] Our ground truth was provided by independent references in art history and Christian iconography and the titles of the artworks analyzed.[4,10] Test paintings depicting saints from Table 1 were scraped from online art databases based on keywords and contextual phrases, as described in Sect. 3. Figure 2 shows examples of keys as well as Saint Peter holding keys. Notice the great variety in sizes, styles, colors, orientations of such keys. Of course the training and test sets were disjunct: we never tested our attribute recognizer on a painting used for its training.[9] We downsampled each image so it matched the pixel (neuron) numbers in the semantic segmentation network. A simple routine computed the pixel location of the center of mass of each figure.

Next, a very simple routine identified the location of the figure nearest to each attribute, according to the figure's visual center of mass. The final step was to assign the saint in the ATTRIBUTE SAINT ASSOCIATION DATABASE in Table 1 to that closest figure. Testing the overall system consisted of presenting a novel Christian painting, classifying each significant actor, and comparing the result to the identification provided by a human expert. In this way we find the percentage of true and false positives, as well as true and false negatives.
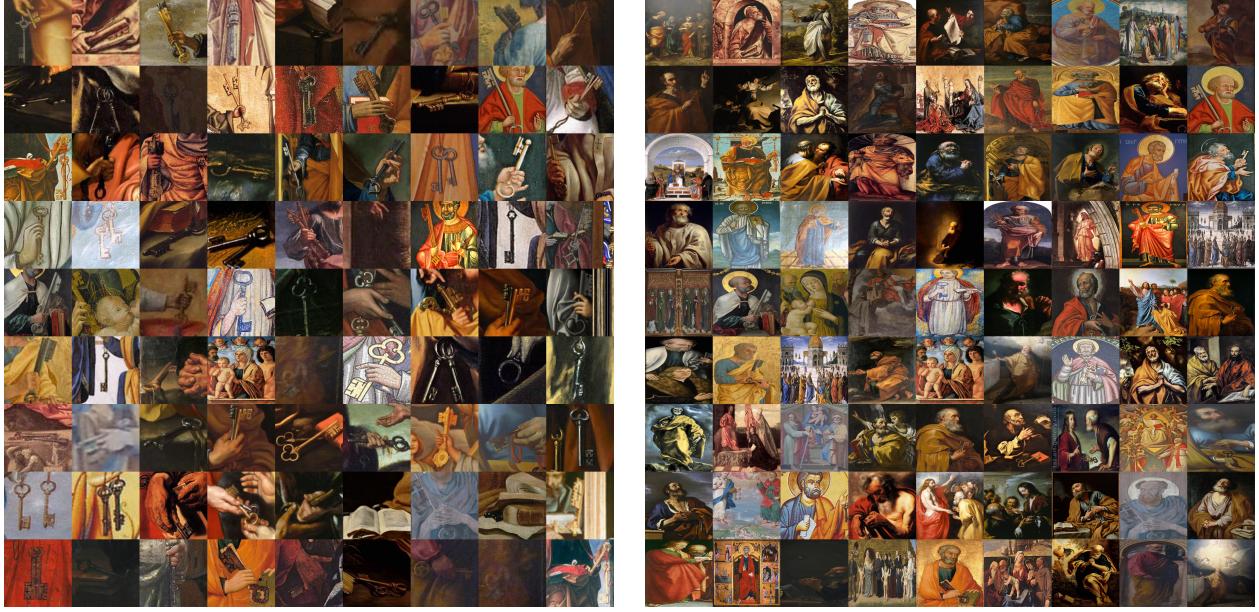
Figure 2. (L) Images of the attribute of keys culled from Western artwork from the 13th through 18th centuries, used for learning of Saint Peter's attribute. (R) Images of Saint Peter, most of which include his attribute.

7. RESULTS AND INTERPRETATIONS

Figure 3 shows the analysis by our trained system on Andrea Verrocchio's *Baptism of Christ.* The left panel shows the output of the ATTRIBUTE RECOGNITION MODULE, which identifies which attributes are most confidently detected—here of DOVE and CROSS—each with an extremely high confidence of 99%. The ATTRIBUTE RECOGNIZER MODULE was applied to each test image to yield the locations and statistical confidences associated that each candidate attribute was present at a given location in the test painting. Only the one or two attributes whose presence was computed to be above a high threshold were retained, which thus yielded a spatial location for that attribute. The network also localizes each such attribute, as indicated by the bounding boxes. The right panel shows the output of the SEMANTIC SEGMENTATION MODULE. Note especially the segmentation regions for PERSON and ANIMAL. Our algorithm merges such regions into a single region, which was then a candidate actor. As is evident from that panel, the semantic segmentation is quite accurate throughout the dataset. We can expect that segmentation will be even more accurate and robust if the network is trained on a larger number of paintings representative of the works in the test set.

Figure 4 shows eight paintings in our dataset and the output of the ATTRIBUTE CLASSIFIER MODULE, specifically two paintings showing the dove, four showing the crucifixion cross, and two showing a lion. Note particularly that the crucifixion cross can be associated with Christ of Saint John (and yet other saints). We then automatically computed the spatial distance between the center of mass of each segmented candidate actor and the location of the confidently identified semiotic attributes. The identity of the actor was assigned that associated with its closest such attribute according to the ATTRIBUTE SAINT DATABASE in Table 1. For example, in *Baptism of Christ*, above, the attributes DOVE and CROSS were associated with Christ and Saint John the Baptist, respectively—a proper "reading" of this painting.

The system performance is expressed as precision and recall, where *precision* represents the proportion of the identifications are actually correct and *recall* represents the proportion actually positives that were correctly identified by the system. Formally, these statistics are defined by:

$$\text{precision} = \frac{TP}{TP + FP} \qquad \text{and} \qquad \text{recall} = \frac{TP}{TP + FN}, \tag{1}$$

5

Figure 3. (L) Andrea Verrocchio's *Baptism of Christ* (177×151 *cm*) oil on wood (1475), Uffizi Gallery, Florence, and bounding boxes of saints' attributes found automatically. Here, the dove and cross are each found with a confidence of 99%. (R) The semantic segmentation where the pink and crimson indicate ANIMAL and PERSON, which jointly define the figures. The proximity of the attributes to the figures properly identify the central figure as Christ and the figure at the right as Saint John the Baptist, the correct reading of this work. Notice the presence of unphysical items, such as halos and golden rays emanating from above.

	God	Mark	John	Peter
Precision	100%	67%	100%	100%
Recall	100%	100%	75%	50%

Table 2. The precision and recall of our system for four saints.

where TP is represents true positives, FP false positives, and FN false negatives. Table 2 shows the precision and recall for our system based on the results in Fig. 4.

8. CONCLUSIONS AND FUTURE DIRECTIONS

Our overarching program is to develop computer-based methods for the interpretation of fine art, in particular to extract rudimentary meanings. Art images differ in many ways from natural photographs, which dominate research into semantic image analysis research. Art explores an extraordinarily larger range of styles than do photographs and unlike most photographs art is frequently crafted by an artist ("author") seeking to express a story, concept, moral, or meaning. As such, it is important that automated system identify not merely the picture's components and their relations (as is common in traditional semantic image analysis), but also what the image *means*.

We have demonstrated that one early step in the interpretation of story or meaning in paintings in the Western canon can be computed automatically. Specifically, our system can reliably identify the actors (here, saints) in Christian art by means of their symbols or attributes. The system relies on special-purpose ATTRIBUTE CLASSIFIER MODULE, existing SEMANTIC SEGMENTATION MODULE networks, and a curated associative ATTRIBUTE SAINT ASSOCIATION DATABASE.
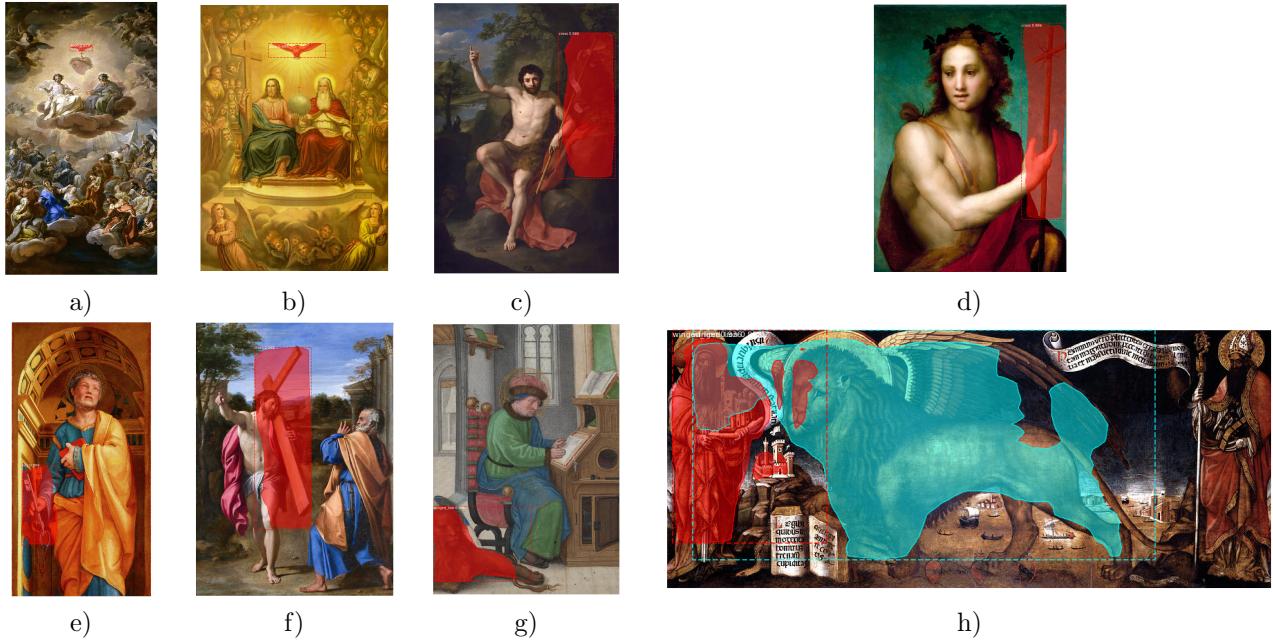
Figure 4. The output of the ATTRIBUTE RECOGNITION MODULE, here accurately finding attributes in the following paintings: a) Corrado Giaquinto's *The holy trinity* (c. 1755), b) unknown artist *Holy trinity* (unknown), c) Anton Raphael Mengs' *St. John the Baptist preaching the wilderness* (1760s), d) Andrea del Sarto's *Saint John the Baptist* (c. 1517), e) Bernardino Zenale's *Saint Peter the apostle* (c. 1510–12), f) Annibale Carracci's *Christ appearing to Saint Peter on the Appian Way* (1601–02), g) Simon Bening's *St. Mark writing* (1521), and h) Donato Veneziano's *The lion "andante" of St. Mark* (1459).

There are a number of complexities and challenges in expanding this work to larger corpora, even if restricted to religious art in the Western canon. First, the total number of saints with attributes depicted in artwork is in the hundreds. Some of these associated attributes are complex and difficult to recognize, such as silk gloves—the attribute of Saint Louis of Toulouse. The number and variety of attributes or symbols is likely equally large. The computational task becomes even more challenging because multiple saints share the same attributes. For those cases, the attribute taken alone is not sufficient for accurate actor identification. A more probabilistic inference may be necessary, one based on some global criterion of all the actors and their simultaneous presence in candidate stories or episodes.

Recall that our ATTRIBUTE SAINT ASSOCIATION DATABASE was hand curated for this preliminary proof-of-concept work. Future work based on natural language analysis of source texts, such as titles of artworks, Biblical source stories, and art historical analyses may be sufficient to learn the associations between saints and attributes. A great benefit of this approach would be to learn concurrence probabilities, which would likely improve the accuracy of the overall system. Of course all results are probabilistic, so this level of analysis may yield two or three candidate saints, each with a confidence score. In such cases, the presence of other actors in the scene—each with its own confidence score—can likely lead to an improved recognition of each *overall*. Although existing segmentation methods were adequate for the work presented here, we can expect improved performance on challenging large datasets if transfer learning is used with representative art images themselves.

ACKNOWLEDGEMENTS

REFERENCES

1. I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, Adaptive computation and machine learning, MIT Press, Cambridge, MA, 2016.
2. P. de Rynck, *How to read a painting: Lessons from the Old Masters*, Harry N. Abrams, New York, NY, 2004.
3. S. Zuffi, *How to read Italian Renaissance painting*, Harry N. Abrams, New York, NY, 2010.
4. G. Ferguson, *Signs & symbols in Christian art*, Oxford University Press, Oxford, UK, 1990.
5. G. Aiello, "Visual semiotics: Key concepts and new directions," in *SAGE handbook of visual research methods*, L. Pauwels and D. Mannay, eds., ch. 23, pp. 367–380, SAGE Publications, Thousand Oaks, CA, 2019.
6. S. Moriarty, "Visual semiotics theory," in *Handbook of visual communication: Theory, methods and media*, K. Smith, S. Moriary, G. Barbatsis, and K. Kenney, eds., ch. 15, pp. 227–241, Lawrence Erlbaum, Mahwah, NJ, 2005.
7. R. Martin, *Classical mythology: The basics*, Routledge, London, UK, 2016.
8. K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *International Conference on Computer Vision (ICCV)*, pp. 2961–2969, October 2017.
9. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, John Wiley and Sons, New York, NY, Second ed., 2001.
10. D. Apostolos-Cappadona, *A guide to Christian art*, T & T Clark Publishers, London, UK, 2020.