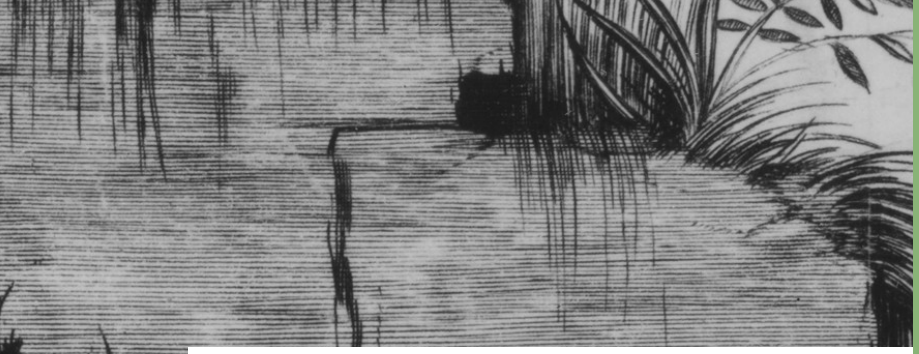
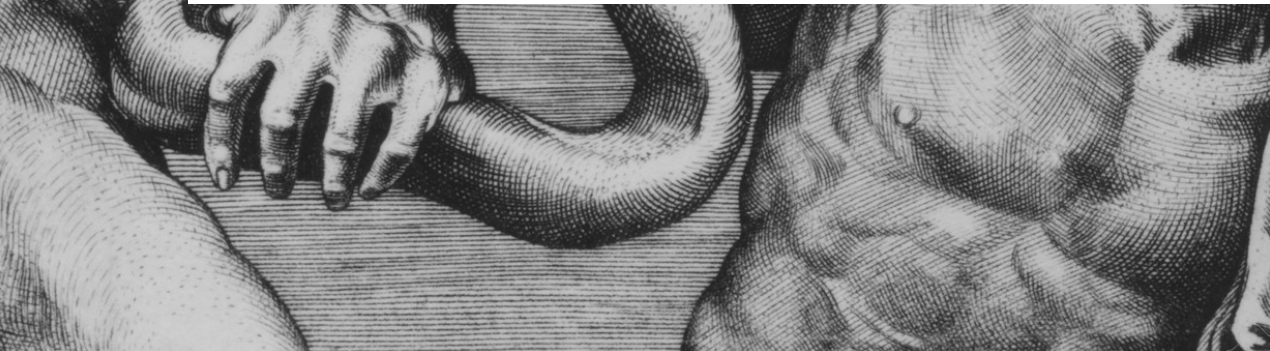


1 introduction

This book explores how computational methodologies, through their 'artificial eye,' create a *virtual space* of associations, references, and similarities, to historically situate the human figure, and its posture, in the visual arts. It focuses on the estimation of joint positions to produce a mathematically tractable vector representation of the human skeleton, which is then utilized to retrieve human figures of potential relevance. This chapter provides the conceptual framework for the subsequent analysis, defining its terminology of *body-related behavior*, *body-related models*, and *body-related similarity*, before introducing the research questions to be addressed in the following chapters.



introduction

The human form is essential to visual, non-verbal communication; it allows emotions, actions, and intentions to be deciphered solely through an intuitive understanding of human physiognomy.[1] Especially in art history, a discipline inherently linked to the visual, the human form is almost ubiquitous, not only as a subject of representation but as a *carrier of meaning*—along with the body movements and postures that emanate from it.[2] Yet art-historical research has shown a predilection for the study of gesture, as a semantically charged sub-category of body movement, over posture.[3] This emphasis can be traced back to the influence of Renaissance humanism, which heralded a re-invigoration of scholarly rhetoric and, thus, gesture[4]—whether it is symbolically performative, as in the medieval Heidelberg *Sachsenspiegel,*[5] or status-identifying, as in

1 Jürgen Ruesch and Weldon Kees (1956). *Nonverbal Communication: Notes on the Visual Perception of Human Relations.* Berkeley, CA: University of California Press.
2 Margreth Egidi et al. (2000). "Riskante Gesten. Einleitung." In: *Gestik. Figuren des Körpers in Text und Bild*. Ed. by Margreth Egidi et al. Literatur und Anthropologie 8. Tübingen: Narr, 11–41, here 30.
3 This is also emphasized by Emmelyn Butterfield-Rosen (2021). *Modern Art and the Remaking of Human Disposition*. Chicago, IL: University of Chicago Press, 21.
4 Michael F. Zimmermann (2011). "Die Sprache der Gesten und der Ursprung der menschlichen Kommunikation. Bildwissenschaftliche Überlegungen im Ausgang von Leonardo." In: *Aufbrüche. Für Andreas Lob-Hüdepohl*. Ed. by Heiner Böttger et al. Eichstätt: Academic Press, 178–197, here 179.
5 Karl von Amira (1905). *Die Handgebärden in den Bilderhandschriften des Sachsenspiegels*. Vol. 23. Abhandlungen der Bayerischen Akademie der Wissenschaften. Philosophisch-Philologische und Historische Klasse 2. Munich: Franz.

Published in: Stefanie Schneider, Art and the Artificial Eye: Algorithmic Views of the Human Posture, Heidelberg: arthistoricum.net, 2025.
https://doi.org/10.11588/arthistoricum.1566

Roman sculpture.[6] Ernst H. Gombrich suggested in 1966 that gestures in the visual arts originate from natural human expressions that have been ritualized, thus acquiring unique forms and meanings.[7] In particular, Aby Warburg's concept of *Pathosformeln* contributed significantly to the discourse by identifying formally stable gestures from antiquity that were repeatedly employed in the Renaissance to express primal emotions.[8] In contrast, through the concept of the "period eye," Michael Baxandall argued that perception is influenced by culturally specific modes of inference; gestures thus reflect the cultural experiences of their time, requiring an examination of these contexts in order to understand their historical significance.[9]

Despite gesture's historiographically proven significance, according to Emmelyn Butterfield-Rosen, "the more useful category" for exploring "permanent, underlying psychic structures" is posture.[10] Even in the early 20th century, the Finnish art historian Johan Jakob Tikkanen was already proposing a semi-consequential typification of leg postures as 'cultural motifs' in the visual arts;[11] these motifs, according to Tikkanen, have evolving functions that may be traced through European art from antiquity to modernity. Despite, or perhaps because of, Tikkanen's

---

6 Richard Brilliant (1963). *Gesture and Rank in Roman Art: The Use of Gestures to Denote Status in Roman Sculpture and Coinage*. New Haven, CT: Connecticut Academy of Arts & Sciences.

7 Ernst H. Gombrich (1966). "Ritualized Gesture and Expression in Art." In: *Philosophical Transactions of the Royal Society of London*. Vol. 251. Series B, Biological Sciences 772, 393–401. Online: <https://doi.org/10.1098/rstb.1966.0025>.

8 See Aby Warburg (1998a). "Der Eintritt des antikisierenden Idealstils in die Malerei der Frührenaissance." In: *Die Erneuerung der heidnischen Antike. Kulturwissenschaftliche Beiträge zur Geschichte der europäischen Renaissance. Gesammelte Schriften*. Ed. by Horst Bredekamp and Michael Diers. Berlin: Akademie Verlag, 173–176; Aby Warburg (1998b). "Dürer und die italienische Antike." In: *Die Erneuerung der heidnischen Antike. Kulturwissenschaftliche Beiträge zur Geschichte der europäischen Renaissance. Gesammelte Schriften*. Ed. by Horst Bredekamp and Michael Diers. Berlin: Akademie Verlag, 443–449; Fritz Saxl (2012). *Gebärde, Form, Ausdruck. Zwei Untersuchungen*. Zürich: Diaphanes.

9 Michael Baxandall (1972). *Painting and Experience in 15th Century Italy: A Primer in the Social History of Pictorial Style*. Oxford: Oxford University Press.

10 Butterfield-Rosen (2021, 21).

11 Johan Jakob Tikkanen (1912). *Die Beinstellungen in der Kunstgeschichte. Ein Beitrag zur Geschichte der künstlerischen Motive*. Helsingfors: Druckerei der finnischen Litteraturgesellschaft.

own admission of the study's limitations and his plea for an interdisciplinary approach to better understand the evolution of motifs,[12] his work remains largely unrecognized in academic discourse, often relegated to the footnotes.[13] Erica Tietze-Conrat, in her review of Tikkanen's study, criticized its minimal focus on what she identified as the "artistic problem"—namely, the factors influencing artists' postural choices. According to Tietze-Conrat, an in-depth analysis of the expressiveness of common figural postures indicates that only a small fraction serves as "vehicles of expression," and those few typically convey a "quite vague and general expressiveness which might only be made clearer when placed in relation to another."[14] Similarly, Desmond Morris's more recent examination of *Posture* (2019) has also been marginalized,[15] partly due to Morris's deliberate avoidance of scientific objectivity and the perceived arbitrariness of the postures analyzed.[16]

This book draws on the previous work of both Tikkanen and Morris; however, it advances from the historically established practice of *close viewing*—i.e., the qualitative analysis of individual artworks within their spatio-temporal contexts—to a quantitative *distant viewing* of

---

12 Ibid., 1.

13 See Peter Schmidt (1992). *Die grosse Schlacht. Ein Historienbild aus der Frühzeit des Kupferstichs*. Wiesbaden: O. Harrassowitz, 45; Silke Meyer (2003). *Die Ikonographie der Nation. Nationalstereotype in der englischen Druckgraphik des 18. Jahrhunderts*. Münster: Waxmann, 127; Leo Steinberg (2018). *Michelangelo's Sculpture: Selected Essays*. Ed. by Sheila Schwartz. Chicago: University of Chicago Press, 200; Butterfield-Rosen (2021), 279.

14 Erica Tietze-Conrat (2020). "On Leg Poses in Art History." Trans. by Karl Johns. In: *Journal of Art Historiography* 23. Ed. by Karl Johns, 1–4, here 2–3. Online: <https://arthistoriography.files.wordpress.com/2020/11/karl-trans-etctikannen-1.pdf> (accessed 17 October 2024). Tikkanen's approach is also questioned by Elke Ullrich (2009). *Das Laszive der Keuschheit in der europäischen Kunst. Die Frau des Potiphar und Joseph von Ägypten. Eine Kulturgeschichte der versuchten Verführung*. Kassel: Kassel University Press, 109.

15 See Butterfield-Rosen (2021, 258); Lisa Hecht (2021). "Degas' gähnende Büglerinnen zwischen Decorumsverstoß und Hysterie." In: *Atem. Gestalterische, ökologische und soziale Dimensionen*. Ed. by Linn Burchert and Iva Rešetar. Berlin: De Gruyter, 31–46, here 33. Online: <https://doi.org/10.1515/9783110701876>.

16 As Morris himself states: "One way to approach the intriguing subject of human body language in art is to take each part of the human body in turn [...] But while this has the merit of being systematic and objective, it is also rather dry and academic" (Desmond Morris [2019]. *Postures: Body Language in Art*. London: Thames & Hudson, 7).

large-scale visual corpora.[17] By harnessing the 'artificial eye' of computational methodologies, it attempts to construct a *virtual space of possibilities*—of associations, references, and similarities—for the historical embedding of the human figure, and its posture, in the visual arts. In this context, it argues for the integration of close and distant viewing, where the global analysis of distant viewing logically precedes and enhances the localized, detailed analysis of close viewing; it does not supplant qualitative analysis of individual objects but, rather, provides a framework for identifying connections between a larger *scale* of objects. From a computational point of view, the approach first *estimates* joint positions of human figures to create a mathematically tractable vector representation of the human skeleton, which is then utilized to *retrieve* figures of potential relevance. This sequence is essential to bridge the *semantic gap*, i.e., the gap between human semantic perception of an image and its computationally measurable features.[18] Methodologically, the book is thus situated within the fields of *Human Pose Estimation* (HPE) and *Human Pose Retrieval* (HPR).

Moreover, it contributes to the expanding field of *Digital Art History* (DAH), which has grown considerably since Johanna Drucker in 2013 criticized the delayed advent of computational methods for art-historical inquiry, despite the increasing availability of suitable data in online repositories.[19] In response, numerous approaches, frameworks, and techniques have emerged over the past decade to computationally examine art-historical research questions;[20] yet the adoption of HPE methods

---

17 For a discussion of the tension between close and distant viewing, see Taylor Arnold and Lauren Tilton (2019). "Distant Viewing: Analyzing Large Visual Corpora." In: *Digital Scholarship in the Humanities* 34, i3–i16. Online: <https://doi.org/10.1093/llc/fqz013>.

18 Arnold W. M. Smeulders et al. (2000). "Content-based Image Retrieval at the End of the Early Years." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.12, 1349–1380, here 1375. Online: <https://doi.org/10.1109/34.895972>.

19 Johanna Drucker (2013). "Is There a 'Digital' Art History?" In: *Visual Resources* 29, 5–13. Online: <https://doi.org/10.1080/01973762.2013.761106>.

20 See, e.g., Fahad Shahbaz Khan et al. (2014). "Painting-91: A Large Scale Database for Computational Painting Categorization." In: *Machine Vision Applications* 25.6, 1385–1397. Online: <https://doi.org/10.1007/s00138-014-0621-6>; Hui Mao et al. (2017). "DeepArt: Learning Joint Representations of Visual Arts." In: *MM '17: The 25th ACM International Conference on Multimedia*. Ed. by Qiong Liu et al. New

in art history has been limited, with the majority of existing approaches relying on supervised learning models that suffer from the absence of specialized, labeled training data pertinent to the domain. Leonardo Impett and Sabine Süsstrunk address this problem by using only manually labeled data—which simultaneously limits the scope of the analysis to the data at hand and undermines the inherent advantage of computational methods, which are primarily intended to be employed on unlabeled data.[21] To expand the utilization of HPE in DAH, multiple research projects apply models trained on real-world photographs without integrating art-historical material, thus limiting their domain-specific applica-

---

York: ACM, 1183–1191. Online: <https://doi.org/10.1145/3123266.3123405>; Lorenzo Baraldi et al. (2018). "Aligning Text and Document Illustrations: Towards Visually Explainable Digital Humanities." In: *24th International Conference on Pattern Recognition, ICPR 2018*. New York: IEEE, 1097–1102. Online: <https://doi.org/10.1109/ICPR.2018.8545064>; Nicolas Gonthier et al. (2018). "Weakly Supervised Object Detection in Artworks." In: *Computer Vision – ECCV 2018 Workshops*. Ed. by Laura Leal-Taixé and Stefan Roth. Vol. 11130. Lecture Notes in Computer Science. Cham: Springer, 692–709. Online: <https://doi.org/10.1007/978-3-030-11012-3_53>; Gjorgji Strezoski and Marcel Worring (2018). "OmniArt: A Large-scale Artistic Benchmark." In: *ACM Transactions on Multimedia Computing, Communications, and Applications* 14.4, 88:1–88:21. Online: <https://doi.org/10.1145/3273022>; Simone Bianco et al. (2019). "Multitask Painting Categorization by Deep Multibranch Neural Network." In: *Expert Systems with Applications* 135, 90–101. Online: <https://doi.org/10.1016/j.eswa.2019.05.036>; Stefanie Schneider et al. (2020). "The Dissimilar in the Similar: An Attribute-guided Approach to the Subject-specific Classification of Art-historical Objects." In: *50. Jahrestagung der Gesellschaft für Informatik, INFORMATIK 2020*. Ed. by Ralf H. Reussner et al. Vol. P-307. LNI. Bonn: GI, 1355–1364. Online: <https://doi.org/10.18420/inf2020_127>; Eva Cetinic (2021). "Iconographic Image Captioning for Artworks." In: *Pattern Recognition: ICPR International Workshops and Challenges*. Ed. by Alberto Del Bimbo et al. Vol. 12663. Lecture Notes in Computer Science. Cham: Springer, 502–516. Online: <https://doi.org/10.1007/978-3-030-68796-0_36>; Nikolai Ufer et al. (2021). "Large-scale Interactive Retrieval in Art Collections Using Multi-style Feature Aggregation." In: *PLOS ONE* 16.11, 1–38. Online: <https://doi.org/10.1371/journal.pone.0259718>; Tillmann Ohm et al. (2023). "Collection Space Navigator: An Interactive Visualization Interface for Multidimensional Datasets." In: *International Symposium on Visual Information Communication and Interaction, VINCI 2023*. Online: <https://doi.org/10.1145/3615522.3615546>; Matthias Springstein et al. (2024). "Visual Narratives: Large-scale Hierarchical Classification of Art-historical Images." In: *IEEE Winter Conference on Applications of Computer Vision, WACV 2024*. New York: IEEE, 7220–7230. Online: <https://doi.org/10.1109/WACV57701.2024.00705>.

21 Leonardo Impett and Sabine Süsstrunk (2016). "Pose and Pathosformel in Aby Warburg's Bilderatlas." In: *Computer Vision – ECCV 2016 Workshops*. Ed. by Gang Hua and Hervé Jégou. Vol. 9913. Lecture Notes in Computer Science. Cham: Springer, 888–902. Online: <https://doi.org/10.1007/978-3-319-46604-0_61>.

tion.[22] In addition, Prathmesh Madhu et al. explore the potential of style transfer to adapt real-world photographs to the aesthetic characteristics of art-historical images, thereby enhancing their utility for art-historical research.[23] They also propose refining the accuracy of pre-trained models through domain-specific fine-tuning.

1 TERMINOLOGY

Before discussing the research questions and contributions of this book, we briefly outline its terminological framework. It consists of three concepts: body-related behavior, body-related models, and body-related similarity.

1.1 BODY-RELATED BEHAVIOR

The attempt to forge a unified terminology for what Doris Schöps calls *body-related behavior*[24] has, at least in the visual arts, only been partial-

---

22 See Tomás Jenícek and Ondrej Chum (2019). "Linking Art Through Human Poses." In: *International Conference on Document Analysis and Recognition, ICDAR 2019*. New York: IEEE, 1338–1345. Online: <https://doi.org/10.1109/ICDAR.2019.00216>; Prathmesh Madhu et al. (2020). "Understanding Compositional Structures in Art Historical Images Using Pose and Gaze Priors: Towards Scene Understanding in Digital Art History." In: *Computer Vision – ECCV 2020 Workshops*. Ed. by Adrien Bartoli and Andrea Fusiello. Vol. 12536. Lecture Notes in Computer Science. Cham: Springer, 109–125. Online: <https://doi.org/10.1007/978-3-030-66096-3_9>; Shu Zhao et al. (2022). "Automatic Analysis of Human Body Representations in Western Art." In: *Computer Vision – ECCV 2022 Workshops*. Ed. by Leonid Karlinsky et al. Vol. 13801. Lecture Notes in Computer Science. Cham: Springer, 282–297. Online: <https://doi.org/10.1007/978-3-031-25056-9_19>.

23 Prathmesh Madhu et al. (2023). "Enhancing Human Pose Estimation in Ancient Vase Paintings via Perceptually-grounded Style Transfer Learning." In: *Journal on Computing and Cultural Heritage* 16.1, 1–17. Online: <https://doi.org/10.1145/3569089>.

24 Doris Schöps (2012). "Semantik und Pragmatik von Körperhaltungen im Spielfilm." In: *IMAGE. Zeitschrift für interdisziplinäre Bildwissenschaft* 8.2, 5–33. Online: <https://doi.org/10.25969/mediarep/16553>.

ly realized.[25] To delineate the existing terminological ambiguity, Butterfield-Rosen distinguishes between gesture, which pertains to localized movements of body parts—especially the head and hands—and posture, which describes the entire body's stance, extending beyond just the upper extremities to include the lower body. Compared to gesture, posture is, according to Butterfield-Rosen, a "lower-order, more primary phenomenon"; it serves as a comprehensive descriptor of an individual's spatial orientation, detailing their "up-down, front-back and left-right orientation," as well as their interaction with the environment.[26] For Carl Sittl, gestures are non-mechanical movements of the human body, divided into categories of instinctive and consciously initiated movements,[27] while Moshe Barasch distinguishes between "symptomatic" gestures, which are natural to human behavior, and "conventional" gestures, which result from social conventions.[28] In contrast, Jean-Claude Schmitt utilizes the Latin *gestus*, which denotes both a general movement—or

---

25 See, e.g., Walter Löffler (1987). "Haltung und Bewegung als Ausdruckssprache bei Michelangelo. Das Vokabular und seine Verwendung." PhD thesis. Eberhard Karls Universität Tübingen; Jan N. Bremmer and Herman Roodenburg, eds. (1992). *A Cultural History of Gesture*. Ithaca, NY: Cornell University Press; Ilsebill Barta Fliedl and Christoph Geissmar (1992). *Die Beredsamkeit des Leibes. Zur Körpersprache in der Kunst*. Salzburg/Vienna: Residenz Verlag; Roland Kanz (1993). *Dichter und Denker im Porträt. Spurengänge zur deutschen Porträtkultur des 18. Jahrhunderts*. Kunstwissenschaftliche Studien 59. Munich: Deutscher Kunstverlag; André Chastel (2001). *Le Geste dans l'art*. Paris: Liana Levi; Clifford Davidson, ed. (2001). *Gesture in Medieval Drama and Art*. Early Drama, Art, and Music Monograph Series 28. Kalamazoo, MI: Medieval Institute Publications; Florens Deuchler (2014). *Strukturen und Schauplätze der Gestik. Gebärden und ihre Handlungsorte in der Malerei des ausgehenden Mittelalters*. Berlin: De Gruyter. In particular, the challenge of defining gesture in a universally accepted manner is evident, as noted by Axel Gampp (2008). "Antikendämmerung. Vom eingeschränkten Gebrauch der Gesten als Bildsprache im frühen Mittelalter." In: *Riten, Gesten, Zeremonien. Gesellschaftliche Symbolik in Mittelalter und Früher Neuzeit*. Ed. by Edgar Bierende, Sven Bretfeld, and Klaus Oschema. Trends in Medieval Philology 14. Berlin: De Gruyter, 3. Online: <https://doi.org/10.1515/9783110210897.1.3>, and Martin Franz Mäntele (2010). "Die Gesten im malerischen und zeichnerischen Werk Raffaels." PhD thesis. Eberhard Karls Universität Tübingen, 20.
26 Butterfield-Rosen (2021, 21).
27 Carl Sittl (1890). *Die Gebärden der Griechen und Römer*. Leipzig: B. G. Teubner, 1.
28 Moshe Barasch (1987). *Giotto and the Language of Gesture*. Cambridge: Cambridge University Press, 3.

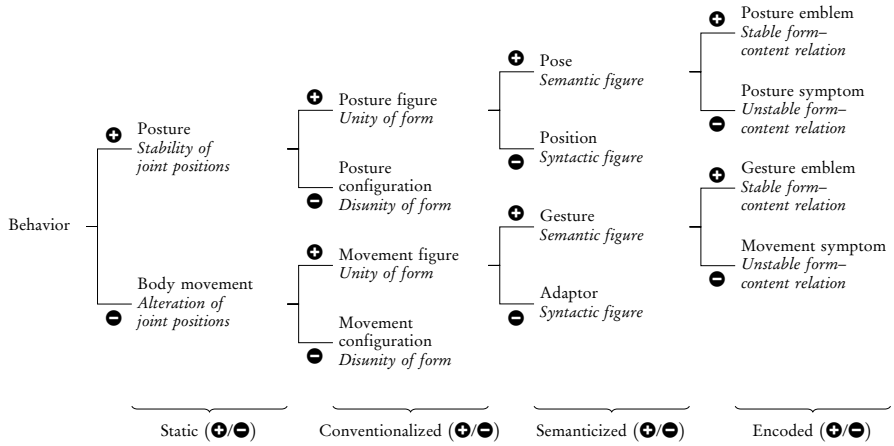


FIG. 1: Taxonomy of body-related behavior using the features *static*, *conventionalized*, *semanticized*, and *encoded*; adapted from Doris Schöps (2012). "Semantik und Pragmatik von Körperhaltungen im Spielfilm." In: *IMAGE. Zeitschrift für interdisziplinäre Bildwissenschaft* 8.2, 5–33, here 7. Online: <https://doi.org/10.25969/mediarep/16553>.

posture—of the body and, more narrowly, the specific movement of a body part, particularly the hand.[29]

To navigate the complex landscape of definitions, we adapt the taxonomy proposed by Schöps,[30] moving from an art-historical to a semiotic framework that stratifies body-related behavior into four hierarchical levels—*static*, *conventionalized*, *semanticized*, and *encoded*—to provide more analytical depth than previous models. As shown in **FIG. 1**, the hierarchy first distinguishes static *postures* from dynamic *body movements*, with postures being described as maintainable physical states and movements as changes in physical state resulting from alterations of joint positions. The second level of the hierarchy classifies body-related behavior into conventionalized *posture* and *movement figures*, which represent a unity of form, and non-conventionalized *posture* and *movement configurations*. In the third level, semantic figures (*poses* and *gestures*) are separated from syntactic figures (*positions* and *adaptors*), with semantic

29 Jean-Claude Schmitt (1992). *Die Logik der Gesten im europäischen Mittelalter*. Stuttgart: Klett-Cotta, 36.
30 Schöps (2012, 7).

ones here also including signs that are either ambiguous or context-dependent. The taxonomy's fourth level considers only encoded figures, *posture* and *gesture emblems*,[31] as having precise, consistent meanings with stable form–content relations, in contrast to non-encoded *posture* and *movement symptoms*.

In this book, we primarily employ the term *posture*, distinguishing between conventionalized posture figures and configurations. We are, however, fully aware that posture, through its figural depiction in the visual arts, can undeniably *convey* movement through the sequential arrangement of narrative scenes or through artificially exaggerated postures; it always captures a moment in which the body is immobile and from which it can be interpreted—a "pregnant" moment that suggests to the viewer "the preceding and following."[32] The term *pose* is used only in cases where a semantic figure is discernible or in conjunction with the application of computational methods in HPE and HPR, where the term is essential for the terminology of the method.

1.2 BODY-RELATED MODELS

As a *model* of the human body, its skeletal framework, consisting of the torso, limbs, and head, provides a sufficient basis for the exploration of body-related behavior.[33] Derived from the Latin *modulus*, a model is a scaled representation of *something* that "may be realized in the physical world."[34] Models embody universally binding laws that govern the relationships, conditions, and structures relevant to the application do-

---

31 See also Paul Ekman and Wallace V. Friesen (1969). "The Repertoire of Nonverbal Behavior: Categories, Origins, Usage and Coding." In: *Semiotica* 1.1, 49–98. Online: <https://doi.org/10.1515/semi.1969.1.1.49>.

32 Gotthold Ephraim Lessing (1766). *Laokoon. Über die Grenzen der Mahlerey und Poesie. Mit beyläufigen Erläuterungen verschiedner Punkte der alten Kunstgeschichte*. Berlin: Christian Friedrich Voss, 26.

33 Pirkko Rathgeber (2019). *Bewegungsfiguren. Über die Bewegung der Strichfigur in der Zeichnung und ihre Bedeutung für den Zeichentrickfilm*. Paderborn: Wilhelm Fink, 87–88.

34 Sanford Friedenthal et al. (2011). *A Practical Guide to SysML: The Systems Modeling Language*. Burlington: Elsevier, 21.

main. They prescribe rules for organizing, quantifying, and defining the array of individual phenomena within their scope;[35] through reduction and abstraction, they distill a *system*—here, the human body—to its essence, simplifying it while preserving important features.[36] Abstraction, as Oskar Schlemmer has put it,

> [...] means simplification, reduction to the essential, to the elementary, to the primary, in order to confront the diversity of things with a unity.[37]

The process of abstraction is "aimed at understanding, communicating, explaining, or designing aspects" of the system under consideration.[38] By abstracting and reducing the human body, we are prompted to inquire about the intrinsic qualities that remain when the body is divested of its psychological and most of its physical identifiers.[39]

The utilization of body-related models is essential in the field of computational analysis because it permits *Convolutional Neural Networks* (CNNs) to process the spatial dimensions of images with human postures. The methodology relies on a training set of abstracted posture representations that guide the neural network in identifying universal

---

35 Ingeborg Reichle et al. (2008). "Die Wirklichkeit visueller Modelle." In: *Visuelle Modelle*. Ed. by Ingeborg Reichle et al. Munich: Wilhelm Fink, 9–16, here 9.

36 Eric Bouvier et al. (1997). "From Crowd Simulation to Airbag Deployment: Particle Systems, a New Paradigm of Simulation." In: *Journal of Electronic Imaging* 6.1, 94–107. Online: <https://doi.org/10.1117/12.261175>.

37 Oskar Schlemmer (1962). "Abstraktion im Tanz." In: *Oskar Schlemmer und die abstrakte Bühne. Die Neue Sammlung München. 20. November 1961 bis 8. Januar 1962*, 25, original: "[Abstraktion] bedeutet die Vereinfachung, die Reduzierung auf das Wesentliche, auf das Elementare, auf das Primäre, um der Vielfalt der Dinge eine Einheit gegenüberzustellen."

38 Dov Dori (2002). *Object-Process Methodology: A Holistic System Paradigm*. New York: Springer, 272.

39 Regarding the moving figure as a model, Rathgeber notes that it is characterized by "abstraction as something 'that disregards reality or given experience,' and by reduction as something that lies in 'simplification,' in reducing to the right measure" (Pirkko Rathgeber [2011]. "Struktur- und Umrissmodelle als schematische Bilder der Bewegung." In: *Rheinsprung 11. Zeitschrift für Bildkritik* 2, 130–164, here 130, original: "Die bewegte Figur als Modell kennzeichnet sich durch Abstraktion als etwas, ‚das von der Wirklichkeit bzw. von der gegebenen Erfahrung absieht' und durch Reduktion, als etwas, das in der ‚Zurückführung', im Reduzieren auf das richtige Mass liegt").

joint and limb patterns; after the training phase, the network's ability to infer human postures from its learned abstractions is evaluated. During training, a so-called *validation set* monitors the model's performance on data it has not seen before, thus facilitating the fine-tuning of hyper-parameters to improve model generalization. The *test set*, utilized after training and validation, serves to determine the model's overall applicability in estimating human posture for unfamiliar data, marking the *inference stage*. The performance on the test set provides an indicator of the model's effectiveness in predicting human postures in previously unseen images.

1.3
BODY-RELATED SIMILARITY

The perception of body-related behavior is modulated by visual, motor, and affective dimensions.[40] Two postures, or body movements, are intuitively perceived as similar if they are variations of a particular action or sequence of actions, despite spatial and temporal variances.[41] However, identifying logically related postures under varying spatial and temporal contexts is not intuitively feasible—nor is capturing shared attributes among different postures within a given posture figure or configuration. The notion of similarity is, as has frequently been debated and critiqued, ubiquitous yet intangible: "everything is like everything and in endless

---

40 Randolph Blake and Maggie Shiffrar (2006). "Perception of Human Motion." In: *Annual Review of Psychology* 58, 47–73. Online: <https://doi.org/10.1146/annurev.psych.57.102904.190152>. See also Truett Allison et al. (2000). "Social Perception from Visual Cues: Role of the STS Region." In: *Trends in Cognitive Sciences* 4.7, 267–278. Online: <https://doi.org/10.1016/s1364-6613(00)01501-1>; Aina Puce and David Perrett (2003). "Electrophysiology and Brain Imaging of Biological Motion." In: *Philosophical Transactions of the Royal Society of London*. Vol. 358. Series B, Biological Sciences 1431, 435–445. Online: <https://doi.org/10.1098/rstb.2002.1221>; Günther Knoblich et al., eds. (2006). *Human Body Perception from the Inside Out*. Advances in Visual Cognition. New York: Oxford University Press.

41 Lucas Kovar and Michael Gleicher (2004). "Automated Extraction and Parameterization of Motions in Large Data Sets." In: *ACM Transactions on Graphics* 23.3, 559–568. Online: <https://doi.org/10.1145/1015706.1015760>.

ways,"[42] as "any two things are similar in some respect or other."[43] Similarity is not an inherent quality of an entity but an attribute designated by observers, contends Nelson Goodman in his diatribe, making it "relative, variable, [and] culture-dependent."[44] Recognizing similarity requires an observer's interpretation and is, thus, inherently shaped by human judgment.[45]

The notion's difficulty might be alleviated if there were an ontology of similarity classes that suggested our perceptions of similarity are directly influenced by the inherent similarities present in nature, as proposed by Don Ross.[46] Such a model would allow postures to be compared based on their adaptability to spatial transformations—translation, rotation, or scaling—even though it might neglect stylistic differences or individual characteristics of the posture's originator. The primary obstacle to comparing postures is that *logically* similar postures may appear *numerically* dissimilar—especially when there are substantial differences in joint orientations and angular velocities—and that numerically similar postures may not necessarily be logically similar, as evidenced by the structural resemblance between reaching for an object and actually touching it.[47] Research that derives similarity scores from joint angles,[48] keypoint co-

---

42 Donald Davidson (1978). "What Metaphors Mean." In: *Critical Inquiry* 5.1, 31–47, here 39.
43 John R. Searle (1981). "Metaphor." In: *Philosophical Perspectives on Metaphor*. Ed. by Mark Johnson. Minneapolis, MN: University of Minnesota Press, 248–285, here 265.
44 Nelson Goodman (1972). "Seven Strictures on Similarity." In: *Problems and Projects*. Ed. by Nelson Goodman. Indianapolis, IN: Bobbs-Merrill, 437.
45 As noted by Michel Foucault: "There are no resemblances without signatures. The world of similarity can only be a world of signs" (Michel Foucault [1966]. *Les mots et les choses. Une archéologie des sciences humaines*. Bibliothèque des Sciences humaines. Paris: Gallimard, 41, original: "Il n'y a pas de ressemblance sans signature. Le monde du similaire ne peut être qu'un monde marqué").
46 Don Ross (1991). "Hume, Resemblance and the Foundations of Psychology." In: *History of Philosophy Quarterly* 8.4, 343–356, here 344.
47 See Kovar and Gleicher (2004).
48 See Keith Grochow et al. (2004). "Style-based Inverse Kinematics." In: *ACM Transactions on Graphics* 23.3, 522–531. Online: <https://doi.org/10.1145/1015706.1015755>; Jacky C. P. Chan et al. (2007). "Immersive Performance Training Tools Using Motion Capture Technology." In: *ICST Conference on Immersive Telecommunications & Workshops, IMMERSCOM 2007*. Ed. by Murat Kunt and Ghassan AlRegib. ICST/ACM. Online: <https://doi.org/10.4108/ICST.IMMERSCOM2007.2102>; Cheng Chen et al. (2011). "Learning a 3D Human Pose Distance Metric from Geometric Pose Descriptor."

ordinates,[49] or geometric features[50] tends to equate numerical similarity with logical similarity, leading to challenges in distinguishing between unrelated postures and variations of the same posture figure or configuration.

2
OBJECTIVES

As previously discussed, strategies for HPE in art-historical images rely heavily on supervised learning models whose performance is limited by the scarcity and homogeneity of available labeled training data—which is either derived from art-historical sources or synthetically generated via style transfer from real-world images. The central research question

---

In: *IEEE Transactions on Visualization and Computer Graphics* 17.11, 1676–1689. Online: <https://doi.org/10.1109/TVCG.2010.272>; Soshi Shimada et al. (2020). "PhysCap: Physically Plausible Monocular 3D Motion Capture in Real Time." In: *ACM Transactions on Graphics* 39.6, 235:1–235:16. Online: <https://doi.org/10.1145/3414685.3417877>; Valerio Marsocci and Lorenzo Lastilla (2021). "POSE-ID-on: A Novel Framework for Artwork Pose Clustering." In: *ISPRS International Journal of Geo-Information* 10.4, Online: <https://doi.org/10.3390/ijgi10040257>.

49 See Lucas Kovar et al. (2002). "Motion Graphs." In: *ACM Transactions on Graphics* 21.3, 473–482. Online: <https://doi.org/10.1145/566654.566605>; Tatsuya Harada et al. (2004). "Quantitative Evaluation Method for Pose and Motion Similarity Based on Human Perception." In: *IEEE/RAS International Conference on Humanoid Robots, Humanoids 2004*. New York: IEEE, 494–512. Online: <https://doi.org/10.1109/ICHR.2004.1442140>; Clifford Kwok-Fung So and George Baciu (2005). "Entropy-based Motion Extraction for Motion Capture Animation." In: *Computer Animation and Virtual Worlds* 16.3–4, 225–235. Online: <https://doi.org/10.1002/cav.107>; Selen Pehlivan and Pinar Duygulu (2011). "A New Pose-based Representation for Recognizing Actions from Multiple Cameras." In: *Computer Vision and Image Understanding* 115.2, 140–151. Online: <https://doi.org/10.1016/j.cviu.2010.11.004>; Impett and Süsstrunk (2016); Jenícek and Chum (2019).

50 See Meinard Müller et al. (2005). "Efficient Content-based Retrieval of Motion Capture Data." In: *ACM Transations on Graphics* 24.3, 677–685. Online: <https://doi.org/10.1145/1073204.1073247>; Edmond S. L. Ho and Taku Komura (2009). "Indexing and Retrieving Motions of Characters in Close Contact." In: *IEEE Transactions on Visualization and Computer Graphics* 15.3, 481–492. Online: <https://doi.org/10.1109/TVCG.2008.199>.

of this book is to investigate the ability of *Semi-supervised Learning* (SSL) models to overcome these challenges:

> *Research Question 1* (RQ 1): "Is it feasible to develop an SSL approach to HPE in art-historical images? What are the advantages, limitations, and challenges of this approach compared to supervised learning techniques?"

Such an SSL approach, however, still requires the compilation of a small yet sufficiently diverse art-historical data set to conduct quantitative, ablation-focused experiments.

A broad spectrum of *low-level approaches* to HPR has been proposed, employing similarity measures based on the spatial coordinates or joint angles of two-dimensional keypoints in real-world images. Nonetheless, these approaches often yield incomplete and inadequate HPR results due to the semantic gap between *logical* similarity perceived by humans and *numerically quantifiable* similarity measures. To mitigate this gap, *high-level approaches* have recently been introduced that exploit the latent layers of neural networks to include perceived relational contexts.[51] However, the suitability of *Deep Learning* (DL) techniques for HPR in art history remains debatable—amplified by the domain's range of human morphology and spatial distortions—resulting in:

> *Research Question 2* (RQ 2): "To what extent are DL models applicable to the retrieval of human fig-

---

51 See Helge Rhodin et al. (2018). "Unsupervised Geometry-aware Representation for 3D Human Pose Estimation." In: *Computer Vision – ECCV 2018 – 15th European Conference*. Ed. by Vittorio Ferrari et al. Vol. 11214. Lecture Notes in Computer Science. Cham: Springer, 765–782. Online: <https://doi.org/10.1007/978-3-030-01249-6_46>; Xuanchi Ren et al. (2020). "Self-supervised Dance Video Synthesis Conditioned on Music." In: *MM '20: The 28th ACM International Conference on Multimedia*. Ed. by Chang Wen Chen et al. New York: ACM, 46–54. Online: <https://doi.org/10.1145/3394171.3413932>; Jennifer J. Sun et al. (2020). "View-invariant Probabilistic Embedding for Human Pose." In: *Computer Vision – ECCV 2020 – 16th European Conference*. Ed. by Andrea Vedaldi et al. Vol. 12350. Lecture Notes in Computer Science. Cham: Springer, 53–70. Online: <https://doi.org/10.1007/978-3-030-58558-7_4>; Ting Liu et al. (2022). "View-invariant, Occlusion-robust Probabilistic Embedding for Human Pose." In: *International Journal of Computer Vision* 130.1, 111–135. Online: <https://doi.org/10.1007/s11263-021-01529-w>.

> ures in art-historical case studies? Is it feasible for these models to function as recommender systems within such a specialized domain?"

In this context, vector representations derived from neural networks have proven beneficial for various art-historical retrieval tasks[52] by capturing semantically pertinent information within dense vector spaces. However, despite the application of *Dimensionality Reduction* (DR) to assess these *embeddings* in Euclidean space, their practical inspection remains underexplored in scholarly discourse. This observation leads to:

> *Research Question 3* (RQ 3): "How can the perceptual space of embeddings, obtained by DL models, be explored and interpreted? What spatial patterns emerge within these spaces, and to what extent can they encode complex art-historical motifs?"

3 CONTRIBUTIONS

The objective of this book is to address the outlined research questions by proposing a semi-supervised, view-invariant approach that facilitates the quantitative systematization of posture in the visual arts and, thus,

---

52 See Ahmed Elgammal et al. (2018). "The Shape of Art History in the Eyes of the Machine." In: *AAAI Conference on Artificial Intelligence, AAAI 2018*. Ed. by Sheila A. McIlraith and Kilian Q. Weinberger. AAAI Press, 2183–2191. Online: <https://doi.org/10.1609/AAAI.V32I1.11894>; Noa Garcia et al. (2019). "Context-aware Embeddings for Automatic Art Analysis." In: *International Conference on Multimedia Retrieval, ICMR 2019*. Ed. by Abdulmotaleb El-Saddik et al. ACM, 25–33. Online: <https://doi.org/10.1145/3323873.3325028>; Nikolai Ufer et al. (2021); Giovanna Castellano and Gennaro Vessio (2022). "A Deep Learning Approach to Clustering Visual Arts." In: *International Journal of Computer Vision* 130.11, 2590–2605. Online: <https://doi.org/10.1007/S11263-022-01664-Y>; Andres Karjus et al. (2023). "Compression Ensembles Quantify Aesthetic Complexity and the Evolution of Visual Art." In: *EPJ Data Science* 12.1. Online: <https://doi.org/10.1140/EPJDS/S13688-023-00397-3>; Fabian Offert and Peter Bell (2024). "imgs.ai: A Multimodal Search Engine for Digital Art History." In: *International Journal for Digital Art History* 9, 5.28–5.39. Online: <https://doi.org/10.11588/dahj.2023.9.91295>; Ohm et al. (2023); Yuguang Zhao et al. (2023). "Zooming In on Style: Exploring Style Perception Using Details of Paintings." In: *International Journal of Computer Vision* 23.2. Online: <https://doi.org/10.1167/jov.23.6.2>.

assists in the detection of body-related behavior essential to the formal renewal of motifs or topoi. Methodologically, the contributions of this book can be divided into two categories.

3.1
HUMAN POSE ESTIMATION

As discussed in **SECTION 2**, current HPE approaches for art-historical images have significant limitations, which are mainly due to the absence of large, domain-specific training data sets. Such limitations motivate the approaches outlined in this book.

Most existing art-historical data sets emphasize formal attributes to classify artists or materials.[53] However, data sets that provide detailed object-level annotation, particularly through bounding boxes, are notably scarce, with only some contributions being the exception.[54] Al-

---

53 See, e.g., Khan et al. (2014); Thomas Mensink and Jan C. van Gemert (2014). "The Rijksmuseum Challenge: Museum-centered Visual Recognition." In: *International Conference on Multimedia Retrieval, ICMR 2014*. Ed. by Mohan S. Kankanhalli et al. New York: ACM, 451–454. Online: <https://doi.org/10.1145/2578726.2578791>; Mao et al. (2017); Strezoski and Worring (2018); Bianco et al. (2019); Matteo Stefanini et al. (2019). "Artpedia: A New Visual-Semantic Dataset with Visual and Contextual Sentences in the Artistic Domain." In: *Image Analysis and Processing – ICIAP 2019 – 20th International Conference*. Ed. by Elisa Ricci et al. Vol. 11752. Lecture Notes in Computer Science. Cham: Springer, 729–740. Online: <https://doi.org/10.1007/978-3-030-30645-8_66>; Nikolaos-Antonios Ypsilantis et al. (2021). "The Met Dataset: Instance-level Recognition for Artworks." In: *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1, NeurIPS Datasets and Benchmarks 2021*. Ed. by Joaquin Vanschoren and Sai-Kit Yeung. Online: <https://datasets-benchmarks-proceedings.neurips.cc/paper/2021/file/5f93f983524def3dca464469d2cf9f3e-Paper-round2.pdf>; Peiyuan Liao et al. (2022). *The ArtBench Dataset: Benchmarking Generative Models with Artworks*. arXiv: 2206.11404.

54 See Pradeep Yarlagadda et al. (2010). "Recognition and Analysis of Objects in Medieval Images." In: *Computer Vision – ECCV 2010 Workshops*. Ed. by Reinhard Koch and Fay Huang. Vol. 6469. Lecture Notes in Computer Science. Cham: Springer, 296–305. Online: <https://doi.org/10.1007/978-3-642-22819-3_30>; Gustavo Carneiro et al. (2012). "Artistic Image Classification: An Analysis on the PRINTART Database." In: *Computer Vision – ECCV 2012*. Ed. by Andrew W. Fitzgibbon et al. Vol. 7575. Lecture Notes in Computer Science. Cham: Springer, 143–157. Online: <https://doi.org/10.1007/978-3-642-33765-9_11>; Impett and Süsstrunk (2016); Nicholas Westlake et al. (2016). "Detecting People in Artwork with CNNs." In: *Computer Vision – ECCV 2016 Workshops*. Vol. 9913. Lecture Notes in Computer Science. Cham:

though Impett and Süsstrunk, along with Xuan Ju et al. and Madhu et al., also identify keypoints for accurately labeling human postures, their data sets remain inaccessible for further scholarly exploration.[55] To address this desideratum, we present an openly licensed data set for HPE, called *Poses of People in Art* (PoPArt), comprising 2,454 images with 10,749 bounding box and 56,154 keypoint annotations. Compared to People-Art,[56] PoPArt offers almost three times as many figure annotations, including a greater number of overlapping figures and small-area instances, establishing it as the most comprehensive data set for art-historical HPE currently available.

While previous research has introduced effective DL models for HPE in art-historical images,[57] the task remains challenging due to variations in human morphology, perceptual distortions, and environmental clutter. This book proposes a novel two-stage teacher–student framework for HPE that integrates labeled and unlabeled data through an SSL approach. In this paradigm, the teacher model—which updates its parameters using the *Exponential Moving Average* (EMA) of the student model's parameters, as described by Antti Tarvainen and Harri Valpola[58]—serves as a pseudo-label generator for bounding box and keypoint annotations: it utilizes weakly augmented unlabeled images for bound-

---

Springer, 825–841. Online: <https://doi.org/10.1007/978-3-319-46604-0_57>; Lorenzo Baraldi et al. (2018). "Aligning Text and Document Illustrations: Towards Visually Explainable Digital Humanities." In: *24th International Conference on Pattern Recognition, ICPR 2018*. New York: IEEE, 1097–1102. Online: <https://doi.org/10.1109/ICPR.2018.8545064>; Gonthier et al. (2018); Mitchell J. P. van Zuijlen et al. (2021). "Materials In Paintings (MIP): An Interdisciplinary Dataset for Perception, Art History, and Computer Vision." In: *PLOS ONE* 16. Online: <https://doi.org/10.1371/journal.pone.0255109>; Artem Reshetnikov et al. (2022). *DEArt: Dataset of European Art*. arXiv: 2211.01226; Xuan Ju et al. (2023). "Human-Art: A Versatile Human-centric Dataset Bridging Natural and Artificial Scenes." In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2023*. New York: IEEE, 618–629. Online: <https://doi.org/10.1109/CVPR52729.2023.00067>; Madhu et al. (2023).

55 See Impett and Süsstrunk (2016); Ju et al. (2023); Madhu et al. (2023).

56 Westlake et al. (2016).

57 See Jenícek and Chum (2019); Madhu et al. (2020); Zhao et al. (2022); Madhu et al. (2023).

58 Antti Tarvainen and Harri Valpola (2017). "Mean Teachers are Better Role Models: Weight-averaged Consistency Targets Improve Semi-supervised Deep Learning Results." In: *5th International Conference on Learning Representations, ICLR 2017*.

ing box detection and similarly augmented cropped bounding boxes for keypoint detection. This methodology is part of a larger class of SSL techniques that incorporate data augmentation to enhance the feedback signals for unlabeled data, or combine pseudo-label generation with consistency enforcement.[59] By employing a transformer-based model, as opposed to a *Region-based Convolutional Neural Network* (R-CNN), our method significantly reduces the number of predicted bounding boxes and keypoints, thereby simplifying the computational process. Experimental results have shown that the proposed approach outperforms models relying predominantly on style transfer by up to 9.3 % in *Average Precision* (AP) for bounding box and 14.8 % for keypoint detection when SSL is integrated.

3.2 HUMAN POSE RETRIEVAL

To facilitate image retrieval using posture figures and configurations that are often difficult to articulate in natural language, this book presents a view-invariant embedding for HPR. The method differs from previous work discussed in **SECTION 1.3** that derives similarity scores using similarity metrics from keyjoint angles, point coordinates, or geometric features. Instead, it employs a Euclidean embedding distance, which is converted to a probability-based score, following the methodology proposed by Ting Liu et al.[60] It addresses the ambiguity associated with projecting three-dimensional postures onto two-dimensional representations, recognizing that postures can be visually identical but

---

59 See David Berthelot et al. (2019). "MixMatch: A Holistic Approach to Semi-supervised Learning." In: *Advances in Neural Information Processing Systems 32, Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019*. Ed. by Hanna M. Wallach et al., 5050–5060; David Berthelot et al. (2020). "ReMixMatch: Semi-supervised Learning with Distribution Matching and Augmentation Anchoring." In: *8th International Conference on Learning Representations, ICLR 2020*; Kihyuk Sohn et al. (2020). "FixMatch: Simplifying Semi-supervised Learning with Consistency and Confidence." In: *Advances in Neural Information Processing Systems 33, Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*. Ed. by Hugo Larochelle.
60 Liu et al. (2022).

mathematically distinct, depending on the observer's point of view. The method's resilience to occlusion, coupled with its streamlined architecture, ensures its adaptability to a wide range of domains without requiring non-standard similarity measures or labor-intensive pre-processing, as showcased using an art-historical subset of Wikidata.[61]

To facilitate the exploration of the embedding space, we also propose a support set of reference postures that enables *Human Pose Classification* (HPC). While remarkable progress has been made in HPC, its applications remain predominantly focused on tangible, real-world situations.[62] Action categories typically focus on basic human actions or are tailored to specific domains,[63] rendering them inappropriate for art-historical contexts that often feature artificial, unconventional, or symbolically charged postures. Existing art-historical methodologies for HPC have

---

61 <https://www.wikidata.org> (accessed 17 October 2024).

62 Maryam Ziaeefard and Robert Bergevin (2015). "Semantic Human Activity Recognition: A Literature Review." In: *Pattern Recognition* 48.8, 2329–2345. Online: <https://doi.org/10.1016/j.patcog.2015.03.006>; Roshan Singh et al. (2019). "Multi-view Recognition System for Human Activity Based on Multiple Features for Video Surveillance System." In: *Multimedia Tools and Applications* 78.12, 17165–17196. Online: <https://doi.org/10.1007/s11042-018-7108-9>; Hongbo Zhang et al. (2019a). "A Comprehensive Survey of Vision-based Human Action Recognition Methods." In: *Sensors* 19.5. Online: <https://doi.org/10.3390/s19051005>; Athanasios Anagnostis et al. (2021). "Human Activity Recognition Through Recurrent Neural Networks for Human–Robot Interaction in Agriculture." In: *Applied Sciences* 11.2188. Online: <https://doi.org/10.3390/app11052188>; MingQi Lu et al. (2020). "Driver Action Recognition Using Deformable and Dilated Faster R-CNN with Optimized Region Proposals." In: *Applied Intelligence* 50.4, 1100–1111. Online: <https://doi.org/10.1007/s10489-019-01603-4>; Jingyuan Liu et al. (2021). "Normalized Human Pose Features for Human Action Video Alignment." In: *IEEE/CVF International Conference on Computer Vision, ICCV 2021*. IEEE, 11501–11511. Online: <https://doi.org/10.1109/ICCV48922.2021.01132>; Yu Kong and Yun Fu (2022). "Human Action Recognition and Prediction: A Survey." In: *International Journal of Computer Vision* 130.5, 1366–1401. Online: <https://doi.org/10.1007/s11263-022-01594-9>; Zehua Sun et al. (2023). "Human Action Recognition from Various Data Modalities: A Review." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.3, 3200–3225. Online: <https://doi.org/10.1109/TPAMI.2022.3183112>.

63 See Manisha Verma et al. (2020). "Yoga-82: A New Dataset for Fine-grained Classification of Human Poses." In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2020*. New York: IEEE, 4472–4479. Online: <https://doi.org/10.1109/CVPRW50498.2020.00527>; Kristina Host and Marina Ivašic-Kos (2022). "An Overview of Human Action Recognition in Sports Based on Computer Vision." In: *Heliyon* 8.6. Online: <https://doi.org/10.1016/j.heliyon.2022.e09633>.

primarily utilized clustering algorithms, such as *k*-means or *k*-medians, which tend to oversimplify or misinterpret the postures' intrinsic meaning.[64] Our method introduces a refined approach by leveraging a support set of 110 art-historical reference postures from the *Iconographic Classification System* (Iconclass)[65] to classify postures across the whole body, upper body, and lower body. Unlike DL models, which necessitate large training data sets, the method requires minimal input, similar to *One-shot Learning* (OSL) paradigms,[66] by using a single example for HPC. The method's effectiveness on art-historical images has been validated by two case studies at the aggregate and individual level.

4 OUTLINE

The structure of the book is as shown in **FIG. 2**. **CHAPTER II** integrates knowledge from the humanities and natural sciences to explore the geometric abstraction of the human body in the sense of an approximation of the human form. **CHAPTER III** introduces, with PoPArt, the first openly licensed data set dedicated to HPE in art-historical images, containing annotations for bounding boxes and keypoints across 22 art-historical styles. The PoPArt data set is leveraged in **CHAPTER IV** in a two-stage HPE approach, which utilizes two transformer models in a semi-supervised teacher–student framework to predict keypoints in human figures, combining both labeled and unlabeled data. In **CHAPTER V**, the transformation of these keypoints into robust, view-invariant posture embeddings for HPR and HPC is detailed, with their utility in

---

64 See Impett and Süsstrunk (2016); Marsocci and Lastilla (2021).
65 Henri van de Waal (1973–1985). *Iconclass: An Iconographic Classification System. Completed and Edited by L. D. Couprie with R. H. Fuchs*. Amsterdam: North-Holland Publishing Company.
66 See Sean Ryan Fanello et al. (2013). "One-shot Learning for Real-time Action Recognition." In: *Iberian Conference on Pattern Recognition and Image Analysis, IbPRIA 2013*. Ed. by João M. Sanches et al. Vol. 7887. Lecture Notes in Computer Science. Cham: Springer, 31–40. Online: <https://doi.org/10.1007/978-3-642-38628-2_4>; Kunyu Peng et al. (2023). "Delving Deep into One-shot Skeleton-based Action Recognition with Diverse Occlusions." In: *IEEE Transactions on Multimedia* 25, 1489–1504. Online: <https://doi.org/10.1109/TMM.2023.3235300>.

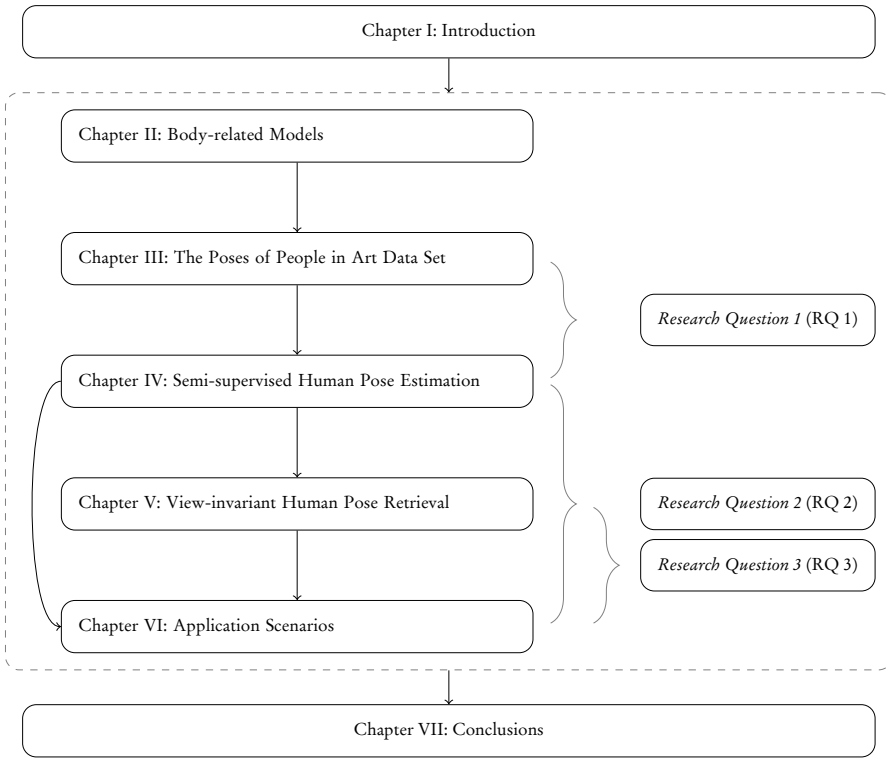


FIG. 2: Book structure with associated research objectives.

art-historical research evaluated by both quantitative and qualitative experiments. The implications of these posture embeddings for art history are considered in **CHAPTER VI** in two case studies. The book concludes with **CHAPTER VII**, summarizing the findings and discussing future avenues of research. Additional details for **CHAPTERS III**, **V**, and **VI** are provided in the **APPENDIX**.