# Evidence for Representational Shifts in the Learning of Chinese Characters: An Investigation using Machine Vision Techniques

**Zhuqian Zhou (zz2404@tc.columbia.edu)[1], James E Corter[1], & Yurong Liang[2]**
[1]Department of Human Development, Teachers College, Columbia University
[2]Bristol Robotics Laboratory, University of the West of England, Bristol

**Abstract**

How does the visual representation of Chinese characters change with expertise? Fluent readers are known to perceive characters in terms of radicals and structural configurations, but the representational basis of perception before such knowledge is acquired remains unclear. We test whether novices' similarity judgments can be explained by generic shape computations from computer vision. Participants with three expertise levels rated pairwise visual similarity of Chinese characters. From pixelated character silhouettes, we computed 109 descriptors spanning multiple families of low-level shape statistics and derived pairwise distances. Using representational similarity analysis and Random Forest prediction, we compared descriptor distances with human judgments. Shape descriptors predicted novices' judgments strongly and learners' moderately but failed to predict fluent readers'. These findings establish a representational bridge between computer vision techniques and human visual cognition, suggesting that the generic shape geometry processing may form the starting point of visual similarity judgments before expertise introduces higher-level structure.

**Keywords:** visual expertise; similarity judgment; computer vision; shape descriptor; representational similarity analysis

## Introduction

Written language is a canonical example of learned visual expertise. Fluent readers recognize words and characters rapidly, robustly, and with high tolerance to variation in size, font, and noise. In cognitive and neural accounts, this proficiency is commonly attributed to the tuning of visual representations in ventral occipito-temporal cortex, including regions often discussed under the visual word form area label, and to the acquisition of efficient orthographic coding that supports lexical access (e.g., Dehaene & Cohen, 2011; Feng et al., 2020). Importantly, learning is not presumed to merely increase sensitivity along fixed perceptual dimensions. Instead, expertise is widely modeled as a process of representational change: observers discover, unitize, and reweight features that are diagnostic for the task, while compressing or discarding irrelevant variation. In reading, these representational changes are expected to reflect the structure of the writing system and the statistical regularities of print.

Logographic writing systems provide a particularly informative testbed for examining this representational transformation. Chinese characters–including Japanese Kanji (Chinese characters used in Japanese) and Korean Hanja (Chinese characters used in Korean), which share many forms) –are visually complex and combinatorial. Many characters can be decomposed into sub-character units such as radicals and logographemes (e.g., 明 bright = 日 sun + 月 moon; 囚 prisoner = 口 cell + 人 person). These units are informative about meaning and pronunciation, and they occur in constrained spatial positions and configurations. Neurocognitive evidence suggests that readers encode radical information in position-specific ways and that sub-lexical composition contributes to neural representations of characters (Liu et al., 2022). Behavioral work likewise indicates that recognition is shaped by recurrent stroke patterns and that multi-component units can function as perceptual units (Isselé et al., 2022). These properties invite a developmental hypothesis: early in learning, observers primarily treat characters as complex shapes and rely on generic visual computations; with experience, observers increasingly encode discrete components and their spatial configurations and may also recruit linguistic information even when tasks are framed as visual.

A complementary way to study representational differences is to examine perceived similarity. Similarity judgments have three advantages here. First, they provide a graded measure of representational proximity rather than a binary outcome. Second, similarity relates directly to orthographic neighborhoods and confusion patterns, linking perception to learning and lexical organization. Third, similarity can be analyzed in a representational similarity framework: if a theory defines a candidate feature space, then distances in that space should preserve the ordering of human similarity judgments to the extent that the model aligns with the observer's representational basis.

Previous studies have demonstrated that similarity measures based on image pixels or image transformation techniques (like Gabor filters) can predict human visual similarity judgments for complex stimuli, including English and Chinese characters, yet these measures account for only a limited proportion of the variance in human data (Zhou & Corter, 2024; Hsiao & Cheung, 2016; Lally & Rastle, 2023). This study addresses this limitation by adopting a comprehensive, computer-vision-derived description of shape. Specifically, we borrow ShapeComp, an image-computable model of shape similarity that aggregates 109



In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* 

complementary descriptors (Morgenstern et al., 2021). ShapeComp was proposed as a generic account of shape similarity that predicts human judgments across diverse shape sets without requiring the fitting of parameters to new human data. From a cognitive modeling perspective, ShapeComp can be interpreted as a broad mid-level feature basis spanning region properties, contour statistics, correspondence-based features, skeleton structure, and Fourier-based representations. If novices' character similarity judgments are grounded in generic visual shape computations, ShapeComp should provide a strong predictor.

A substantial body of research has established that fluent readers do not perceive Chinese characters as undifferentiated shapes. Instead, their perception is organized around meaningful sub-character units, such as radicals and spatial configurations, and is sensitive to the positional legality and combinatorial structure of these components (Zhou & Corter, 2024; Anderson et al., 2013; Yeh & Li, 2002; Yeh et al., 2003). These findings have been central to understanding expert character perception. However, they leave an important developmental question open: what is the representational basis of character perception before such orthographic knowledge is acquired? That is, when observers know that the stimuli are language symbols but lack knowledge of their internal structure, how are visual similarity judgments computed? The present study addresses this gap by asking whether novices' similarity judgments can be explained by the same generic shape computations that support human similarity judgments for ordinary objects and silhouettes in computer vision models.

This approach yields two predictions. First, novices' similarity judgments should be captured well by ShapeComp because novices lack access to orthographic structure and therefore rely heavily on appearance-based cues. Second, the correlation between ShapeComp distance and perceived similarity should weaken with proficiency: learners may show intermediate reliance on generic shape, while fluent readers' judgments may be dominated by structural and linguistic factors beyond silhouette-level shape. By evaluating these predictions on a large multi-proficiency dataset, the present study contributes (i) a quantitative estimate of how much generic shape similarity explains perceived similarity across proficiency levels, (ii) a practical pipeline for importing a comprehensive feature set from computer vision into cognitive similarity analysis, and (iii) constraints on theories of feature learning and representational change in reading.

# Method

## Dataset and Participants

We reanalyzed a subset of similarity judgments originally collected by Yencken and Baldwin (2006) to model orthographic neighborhoods for Japanese Kanji (i.e., Chinese characters in Japanese). The data analyzed here contain 3,182 visual similarity ratings for 61 Kanji pairs drawn from a pool of 165 unique Kanjis. Ratings were provided by 236 participants spanning three proficiency groups: novices (n = 114) with no prior knowledge of Chinese or Japanese who rated 21 Kanji pairs, learners (n = 88) with some knowledge of Chinese or Japanese (but not fluent) who rated all 61 pairs, and fluent readers (n = 34) who rated 36 pairs. Each trial consisted of a Kanji pair and a similarity rating on a five-point scale from 0 (not similar) to 4 (very similar). The data include fields such as participantId, kanjiA, kanjiB, and rating value. We treat the group-mean rating for each pair as the estimate of that group's perceived similarity.

It is worth noting that our study is based on a specific set of Japanese Kanji (Chinese characters in Japanese). While the underlying cognitive mechanisms of visual expertise are expected to generalize to Chinese characters, script-specific factors—such as the visual density of traditional versus simplified characters, and cross-linguistic differences in radical usage frequency between Japanese and Chinese—might influence the exact feature weightings from the analysis described below.

## Stimuli, Image Preprocessing, and Descriptor Construction

To compute image-based descriptors, each of the 165 characters was rendered as a high-resolution binary image using the Microsoft YaHei font at size 900. Using a single font controls typographic variation and focuses the analysis on character form; we revisit the implications of this choice in the Discussion. Rendering proceeded by drawing each character onto a uniform background, converting to grayscale, and thresholding to obtain a binary silhouette.

To operationalize visual shape similarity independent of linguistic knowledge, we adopt the 109 descriptors from ShapeComp (Morgenstern et al., 2021), which were designed to be a parameter-free model to capture human judgments of shape similarity. The descriptors span multiple families that capture complementary aspects of shape (see a detailed explanation of each descriptor from the ShapeComp supporting information here[1]):

Image-based simple shape descriptors (1-22[2]) include standard region properties (area, perimeter, eccentricity, major-axis orientation, major/minor axis ratio, extent, compactness, solidity), and circularity measures A and B (*Circularity A = 4 π (Area/Perimeter²)*; *Circularity B = Perimeter² /Area*), convexity (convex-hull perimeter ratio), centroids (x/y), and Area/Perimeter. In R, these quantities were computed with the *EBImage* and *imager* packages and geometric routines. The intensity-projection statistics along rows and columns (SD, skewness, kurtosis) were computed using R packages, *e1071* and *moments*.

Point-based simple shape descriptors include curviness (23-26[3]), which is the root mean square (RMS) error between

[1] https://doi.org/10.1371/journal.pcbi.1008981.s004

[2] See Footnote 1.

[3] See Footnote 1.

the original boundary and a blurred version of itself at four scales (kernel sizes 2, 4, 8, 16 points). It was implemented by smoothing the contour using the *gblur* function in the R package, *EBImage*, and measuring point-wise deviation. Two other point-based simple descriptors, circular variance (mean-square error to a same-area circle) and elliptical variance (to a same-area ellipse), were also computed (27-28[4]).

For the boundary of each character, shape contexts were computed as a joint histogram of log-polar pairwise distances and angles with 15 distance bins × 15 angle bins (29[5]), and histograms of chord lengths and chord angles with 100 bins each (30-31[6]); shape signatures were computed along arclength, including radius (centroid-to-boundary distance), curvature, triangular area, tangent angle, cumulative tangent angle, and the horizontal/vertical distributions of boundary points (32-38[7]).

Elliptic Fourier descriptors of the closed contour capture the global contour by decomposing the boundary into sinusoidal components. They were computed using the R package, *parcma*, and retained the first ten coefficients as separate descriptors (39-48[8]), plus the complete Fourier description (49[9]), following Kuhl & Giardina's (1982) formulation.

The surprisal descriptors were computed as signed and unsigned information in Shannon's sense along the boundary of a character following Feldman & Singh's (2005) definition (50-51[10]). In this framework, surprisal quantifies how unexpected a local change in contour direction (turning angle, or equivalently curvature) is, given a prior distribution of likely contour continuations. For unsigned information, surprisal depends only on the magnitude of curvature—regions of greater bend, whether convex or concave, carry higher information content because they deviate more strongly from the expected straight continuation of the contour. Signed information extends this analysis to closed contours by incorporating the direction of curvature: positive for convex turns toward the figure and negative for concave turns away from the figure.

For each histogram/signature family above—chord lengths, chord angles, radius, curvature, triangular area, tangent angle, cumulative tangent angle, and the x/y point distributions—mean, standard deviation, skewness, and kurtosis were calculated to yield boundary moments (52-87[11]).

For each one-dimensional signature (32-38[12]), the magnitude of the one-sided Fourier spectrum of it was taken as a descriptor (103-109[13]), computed using the *fft* function in the R package, *Stats*.

The only non-R step was skeletal analysis. For each binary image of Chinese characters the Blum-like medial-axis skeleton was computed using ShapeToolbox in MATLAB, then the skeleton was summarized with 15 descriptors (88-102[14]): total number of branches, maximum and mean skeletal depth, mean branch angle, mean distance along the parent axis at which children emanate, mean axis length relative to the root, integrated absolute unsigned turning angle per axis, integrated absolute signed turning angle per axis, the standard deviations of the angle/distance/length measures, and the distribution of ribs (mean, SD, skewness, kurtosis).

## Descriptor-Based Distances

Each of the 109 descriptors above is either a scalar or a vector. The predicted visual similarity of a pair of Chinese characters based on a particular descriptor was computed as the Euclidean distance between them on that (unidimensional or multidimensional) descriptor *k*:

$$d_k^{ij} = \sqrt{(f_k^i - f_k^j)^2}$$

where $d_k^{ij}$ is the distance between Chinese characters *i* and *j* on shape descriptor *k*, and $f_k^i$ and $f_k^j$ are the values on shape descriptor *k* for Chinese characters *i* and *j*. Once the computation for all pairwise comparisons of the 165 Chinese characters based on a descriptor *k* was complete, the derived 165 × 165 distance matrix was normalized by the largest distance on that descriptor:

$$\hat{d}_k^{ij} = \frac{d_k^{ij}}{\max_k (d_k^{ij})}$$

The normalized distance, $\hat{d}$, was computed for all 109 descriptors to form a 165 × 165 × 109 matrix. Then, an Euclidean distance, *D*, across the descriptors for Chinese characters *i* and *j*, was computed as follows:

$$D^{ij} = \sqrt{\sum_{k=1}^{109} (\hat{d}_k^{ij})^2}$$

## Correlation and Predictive Modeling

To assess the global alignment between the generic shape geometry of Chinese characters and human perception, we computed the Pearson correlation coefficient (*r*) between the aggregate ShapeComp distance—*D*—as described above and the group-mean similarity ratings for each Kanji pair. This simple measure was also used in Morgenstern et al.'s (2021) paper, which allows us to compare the magnitude of the correlations to different sets of visual stimuli. Because our ShapeComp metric calculates geometric distance, higher values indicate greater dissimilarity. Consequently, a strong alignment between the model and human similarity ratings manifests as a large negative correlation.

[4] See Footnote 1.
[5] See Footnote 1.
[6] See Footnote 1.
[7] See Footnote 1.
[8] See Footnote 1.
[9] See Footnote 1.
[10] See Footnote 1.
[11] See Footnote 1.
[12] See Footnote 1.
[13] See Footnote 1.
[14] See Footnote 1.

Because the three proficiency groups rated different subsets of Kanji pairs (21 by novices, 61 by learners, and 36 by fluent readers, as described above), with only partial overlap across groups (16 shared pairs), we conducted not only correlations using all ratings within each proficiency group, but also correlations based exclusively on the 16 pairs rated by all three groups to rule out the effect of different Kanji exposure on *r*. To avoid over-interpreting minor differences in *r* given the relatively small number of Kanji pairs being rated, our interpretation primarily focus on the broad trajectory of these correlations across the three proficiency groups.

Although the characteristic of the ShapeComp model and its aggregated *D* measure is that they do not require any parameter fitting, we still implement a predictive model, Random Forest (RF), to explore whether the model fit can be further improved by optimizing descriptor combinations.

RF is an ensemble learning method that constructs a multitude of decision trees using bootstrap aggregation (bagging). For each split in a tree, a random subset of features is considered, which prevents the model from overfitting to individual dominant predictors (Breiman, 2001). RF is well-suited for this task as it can capture complex geometric configurations that linear models might overlook.

### Evaluation and Feature Importance

Due to the small sample sizes, the RF model was evaluated using Leave-One-Out Cross-Validation (LOOCV). Predictive performance was measured by the Pearson correlation between cross-validated predictions and observed mean ratings, as well as Mean Absolute Error (MAE).

To identify the most diagnostic visual features, we also extracted feature importance metrics. For RF, importance was calculated via Mean Decrease in Impurity (MDI), which measures the total reduction in the sum of squared residuals brought by each feature across all trees in the forest. The larger the MDI, the more influential the feature is.

## Results

### Correlation between Aggregated *D* and Similarity

Pearson correlations between the parameter-free aggregate ShapeComp distance *D* and similarity ratings show that Novice judgments are extremely well-captured by the unweighted shape distance ($r = -.87$, 95% CI [−.95, −.70], $p < .0001$ for all 21 pairs; $r = -.88$, 95% CI [−.96, −.68], $p < .0001$ for 16 shared pairs). This relationship weakens for Learners ($r = -.51$, 95% CI [−.67, −.30], $p < .0001$ for all 61 pairs; $r = -.75$, 95% CI [−.91, −.41], $p < .001$ for 16 shared pairs) and becomes non-significant for Fluent Readers ($r = -.18$, 95% CI [−.48, .15], $p = .29$ for all 36 pairs; $r = -.21$, 95% CI [−.64, .32], $p = .43$ for 16 shared pairs), suggesting that experts discard the global shape similarity in favor of other features.

The correlation between *D* and Novice ratings of Chinese character similarity is comparable to the correlations reported by Morgenstern et al. (2021) between *D* and human similarity ratings of animal silhouettes ($r = -.91$, $p < .01$) and arbitrary shapes ($r = -.78$, $p < .01$).

### Predictive Modeling Performance

Results from the predictive modeling reflect an analogous proficiency trend. For novices, RF models maintained high alignment with human ratings ($r = .63$, 95% CI [0.26, 0.84], $p < .005$, MAE = 0.62). This pattern persists in the learner group, with RF achieving a correlation of $r = .55$, 95% CI [0.35, 0.70], $p < .0001$) and a relatively low error (MAE = 0.38). Crucially, the predictive failure for experts persists even when allowing for optimized feature weighting and non-linear interactions ($r = -.24$, 95% CI [−.53, .10], $p = .16$, MAE = 1.17), further reinforcing the conclusion that experts utilize representations not captured by silhouette-based geometry.

### Feature Importance Analysis

To identify the most diagnostic shape features, we analyzed the MDI from the model. Ordering MDI from the largest to lowest, Table 1 shows that novice similarity judgments are primarily driven by contour statistics and global geometric context. This includes the character's overall footprint (Extent), mass balance (Centroid B), and degree of roundness versus elongation (Circular Variance). The presence of Fourier descriptors suggests that novices are sensitive to the periodic and distributional patterns of ink on the page, rather than specific strokes or radicals. Effectively, novices perceive Chinese characters as generic geometric entities with distinct spatial envelopes and coarse structural complexities.

Table 1: Top five feature importance for modeling novices' ratings according to MDI.

| Descriptor (Feature) | Description | MDI |
|---|---|---|
| Mean of chord angles | Average angle between points on the contour, summarizing internal geometric turns | 0.089 |
| Extent | Ratio of shape pixels to the total bounding box, indicating fill-density | 0.060 |
| Centroid B | Vertical coordinate (Y) of the shape's center of mass | 0.058 |
| Fourier transform of distribution of X values | Frequency-based decomposition of the horizontal point distribution, capturing periodic structure | 0.058 |
| Circular variance | Deviation from a circle of equal area, reflecting overall circularity | 0.056 |

While novices' judgments rely heavily on global region statistics (e.g., extent), learners show a distinct shift in feature weighting (see Table 2). Specifically, Shape Context—a correspondence-based descriptor that captures the relative

spatial configuration of all points on a contour—become the most important predictor for learners (MDI = 0.271) and much more important than it is for novices (MDI = 0.030), which indicates that learners seem to notice the layout and configuration of Kanji while novices mainly see a global blob of ink.

Table 2: Top five feature importance for modeling learners' ratings according to MDI.

| Descriptor (Feature) | Description | MDI |
|---|---|---|
| Shape context | A joint distribution of distance and angle from a starting point connecting to all the rest of points on the contour, capturing local spatial arrangement | 0.271 |
| Skewness for histogram of chord angles | Asymmetry of the distribution of chord angles, capturing directional imbalance | 0.119 |
| Mean for histogram of chord angles | Average value of the distribution of chord angles, capturing the dominant global orientation of the shape | 0.065 |
| Fourier decomposition, harmonic coefficient = 1 | The first low-frequency coefficient obtained by Fourier transforming a contour, capturing the dominant global shape structure | 0.058 |
| Centroid A | Horizontal coordinate of shape's center of mass | 0.045 |

# Discussion

## Summary of Findings

This study investigated whether a structured set of shape descriptors—originally developed in computer vision and formalized in ShapeComp—can explain human judgments of visual similarity for logographic characters across proficiency levels. Two results stand out.

First, ShapeComp distance strongly tracked novice similarity judgments, was moderately aligned for learners, and was non-predictive for fluent readers. This same gradient appeared in the predictive RF model: cross-validated performance was high for novices, moderate for learners, and insignificant for fluent readers.

Second, this novice-learner-fluent drop is not a mere artifact of a single analysis choice: it appears both in (i) correlation between perceived similarity and ShapeComp distances and (ii) model-based prediction of perceived similarity from descriptor vectors. Together, these patterns support a principled interpretation: generic shape geometry appears to constitute the initial representational basis for similarity judgments, which is progressively replaced—not merely reweighted—by orthographic and structural coding as expertise develops.

The present results can be understood most clearly when viewed against prior findings on expert character perception. Numerous studies have shown that fluent readers organize characters in terms of radicals, structural configurations, and position-sensitive component representations (Zhou & Corter, 2024; Anderson et al., 2013; Yeh & Li, 2002; Yeh et al., 2003). In these accounts, similarity between characters is driven by overlap in meaningful orthographic units rather than by global shape resemblance. What has remained largely unaddressed is the complementary question: how similarity is computed *before* such orthographic structure is available. Our findings suggest that, at this early stage, observers rely on a representational basis that closely resembles the generic multidimensional shape space formalized in computer vision. In other words, novices appear to perceive Chinese characters much like they perceive arbitrary shapes (Morgenstern et al., 2021).

## Educational Implications: Visual Expertise for Learning Characters

What do these results imply about learning mechanisms? A purely attentional reweighting account predicts that all groups rely on the same underlying feature basis but differ in weights. The novice-to-learner transition may be consistent with reweighting within a broad visual feature space, because the correlation decreases rather than collapses. However, the learner-to-fluent transition appears more qualitative: the near-zero association even with feature-weighted predictive models suggests that fluent similarity judgments are not simply a different weighting of generic silhouette features. Instead, expertise likely introduces additional representational ingredients—such as component identities and combinatorial constraints—that are not present in a generic shape space. This interpretation aligns with hierarchical accounts of reading in which early processing exploits general visual computations and later processing incorporates orthographic structure learned from print statistics and task demands (Feng et al., 2020).

An educational implication of this staged account is that visual skill in character learning is unlikely to be unitary. Although perceptual expertise measures for Chinese characters have been shown to predict reading performance and to differentiate typical readers from children with developmental dyslexia (Wong et al., 2021), our findings add nuance to this picture. Our findings suggest that, early in learning, visual discrimination may be well approximated by generic multidimensional shape computations, but as learners become fluent, successful reading and successful similarity judgments increasingly depend on structural coding and the integration of orthographic units with language.

In other domains, perceptual learning modules (PLMs) have successfully accelerated the acquisition of expert-like pattern recognition through dense, feedback-guided exposure to informative stimulus variation (Kellman & Massey, 2013; Kellman et al., 2010). The present results suggest that similar

approaches—designed to guide learners from reliance on global shape cues toward sensitivity to orthographic structure—may offer a promising direction for supporting Chinese character learning.

### Methodological Implications: A Representational Bridge between Computation and Cognition

ShapeComp operationalizes shape similarity using a large panel of descriptors spanning boundary geometry, global moments, skeletal structure, convexity/solidity, and Fourier-based contour summaries, among others (Morgenstern et al., 2021). The strong novice alignment indicates that human novices' similarity space is well approximated by a combination of broadly vision-generic shape statistics—the type of representational basis routinely used in computer vision and computational shape analysis.

Notably, several of the descriptor families included in ShapeComp—such as medial-axis skeletons (Wilder et al., 2001) and curvature-based contour information and boundary surprisal (Feldman & Singh, 2005)—were originally proposed in cognitive and psychological theories of shape representation before being formalized and widely adopted in computer vision, creating a historical continuity between cognitive accounts of shape and their computational implementations.

Meanwhile, ShapeComp was never designed to model symbol perception. Its success in predicting novices' similarity judgments of Chinese characters indicates that early character perception operates in a visual feature space shared with general object perception. This cross-domain applicability is precisely what allows computer vision descriptors to serve as a bridge between visual computation and cognitive representation.

### Limitations and Future Directions

Several limitations constrain the present conclusions and motivate follow-up work.

In the original dataset, three proficiency groups were randomly assigned with likely different character pairs to rate, affecting the stability and comparability of correlations. The number of pairs is also limited. A strong next study would use the same and larger pair set across all groups (or a balanced subset) to isolate proficiency effects and to estimate the full similarity space.

The clearest theoretical next step is to build hybrid models that combine ShapeComp with explicit orthographic structure, such as radical identity, radical position, and configuration templates. Such models could directly test whether fluent judgments are predicted by structural overlap and whether learners' judgments reflect a mixture of shape and structure. Existing evidence for position-specific radical coding and multi-component perceptual units provides a principled starting point for specifying such structural features (Isselé et al., 2022; Liu et al., 2022).

It is worth noting that phonological and semantic features of characters should be added to future models in case fluent readers involuntarily activate these associations even in nominally visual tasks. With these control variables, the models could test whether fluent readers' character shape visual representations are truly overwritten, or merely overshadowed by automatic linguistic associations.

## Conclusion

By importing a broad, image-computable model of shape measures into the study of perceptual similarity and logographic orthography, this work strengthens the bridge between computer-vision descriptors and cognitive representations of written symbols. ShapeComp distances strongly predict novices' visual similarity judgments and moderately predict learners' judgments, indicating that developing readers rely on rich multidimensional shape information beyond simple global metrics. Fluent readers' judgments, however, are not captured by the same shape representation. Together, these findings provide direct evidence for a representational shift in the learning of Chinese characters. Fluent readers' similarity judgments reflect orthographic units and structural knowledge documented in prior research, whereas novices' judgments are well explained by generic shape computations used in computer vision and object perception. Learning, therefore, does not simply refine sensitivity within a fixed visual space; it qualitatively transforms the representational basis on which similarity is computed.

## References


Zhou, Z., & Corter, J. E. (2024). Visual Similarity Modeling of Chinese Characters Across Natives, Second Language Learners, and Novices. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 46, pp. 3478-3484).

Anderson, R. C., Ku, Y. M., Li, W., Chen, X., Wu, X., & Shu, H. (2013). Learning to see the patterns in Chinese characters. *Scientific Studies of Reading*, *17*(1), 41-56. https://doi.org/10.1080/10888438.2012.689789

Belongie, & Malik. (2000, June). Matching with shape contexts. In *2000 Proceedings workshop on content-based access of image and video libraries* (pp. 20-26). IEEE. https://doi.org/10.1109/IVL.2000.853834

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in cognitive sciences,* *15*(6), 254-262. https://doi.org/10.1016/j.tics.2011.04.003

Feldman, J., & Singh, M. (2005). Information along contours and object boundaries. *Psychological review*, *112*(1), 243. https://doi.org/10.1037/0033-295x.112.1.243

Feng, X., Altarelli, I., Monzalvo, K., Ding, G., Ramus, F., Shu, H., Dehaene, S., Meng, X., & Dehaene-Lambertz, G. (2020). A universal reading network and its modulation by writing system and reading ability in French and Chinese children. *eLife,* *9*, e54591. https://doi.org/10.7554/eLife.54591

Hsiao, J. H., & Cheung, K. (2016). Visual similarity of words alone can modulate hemispheric lateralization in visual word recognition: Evidence from modeling Chinese character recognition. *Cognitive Science, 40*(2), 351-372. https://doi.org/10.1111/cogs.12233

Isselé, J., Chetail, F., & Content, A. (2022). The nature of perceptual units in Chinese character recognition. *Quarterly Journal of Experimental Psychology*, *75*(8), 1514-1527. https://doi.org/10.1177/17470218211056895

Kellman, P. J., & Massey, C. M. (2013). Perceptual learning, cognition, and expertise. In *Psychology of learning and motivation* (Vol. 58, pp. 117-165). Academic Press. https://doi.org/10.1016/B978-0-12-407237-4.00004-9

Kellman, P. J., Massey, C. M., & Son, J. (2010). Perceptual learning modules in mathematics: enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science (Special Issue on Perceptual Learning)*, *2*(2), 285–305. https://doi.org/10.1111/j.1756-8765.2009.01053.x

Kuhl, F. P., & Giardina, C. R. (1982). Elliptic Fourier features of a closed contour. *Computer graphics and image processing*, *18*(3), 236-258. https://doi.org/10.1016/0146-664X(82)90034-X

Lally, C., & Rastle, K. (2023). Orthographic and feature-level contributions to letter identification. *Quarterly Journal of Experimental Psychology*, *76*(5), 1111-1119. https://doi.org/10.1177/17470218221106155

Liu, X., Wisniewski, D., Vermeylen, L., Palenciano, A. F., Liu, W., & Brysbaert, M. (2022). The representations of Chinese characters: evidence from sublexical components. *Journal of Neuroscience*, *42*(1), 135-144. https://doi.org/10.1523/JNEUROSCI.1057-21.2021

Morgenstern, Y., Hartmann, F., Schmidt, F., Tiedemann, H., Prokott, E., Maiello, G., & Fleming, R. W. (2021). An image-computable model of human visual shape similarity. *PLoS computational biology*, *17*(6), e1008981. https://doi.org/10.1371/journal.pcbi.1008981

Paulun, V. C., Kawabe, T., Nishida, S. Y., & Fleming, R. W. (2015). Seeing liquids from static snapshots. *Vision research*, *115*, 163-174. https://doi.org/10.1016/j.visres.2015.01.023

Peura, M., & Iivarinen, J. (1997, May). Efficiency of simple shape descriptors. In *Proceedings of the third international workshop on visual form* (Vol. 5, pp. 443-451).

Wilder, J., Feldman, J., & Singh, M. (2011). Superordinate shape classification using natural shape statistics. *Cognition*, *119*(3), 325-340. https://doi.org/10.1016/j.cognition.2011.01.009

Wong, Y. K., Tong, C. K. Y., Lui, M., & Wong, A. C. N. (2021). Perceptual expertise with Chinese characters predicts Chinese reading performance among Hong Kong Chinese children with developmental dyslexia. *Plos one*, *16*(1), e0243440. https://doi.org/10.1371/journal.pone.0243440

Yeh, S. L., & Li, J. L. (2002). Role of structure and component in judgments of visual similarity of Chinese characters. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 933. https://doi.org/10.1037/0096-1523.28.4.933

Yeh, S. L., Li, J. L., Takeuchi, T., Sun, V., & Liu, W. R. (2003). The role of learning experience on the perceptual organization of Chinese characters. *Visual Cognition*, *10*(6), 729-764. https://doi.org/10.1080/13506280344000077

Yencken, L., & Baldwin, T. (2006, December). Modelling the orthographic neighbourhood for Japanese kanji. In *International Conference on Computer Processing of Oriental Languages* (pp. 321-332). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/11940098_33

Zhang, D., & Lu, G. (2004). Review of shape representation and description techniques. *Pattern recognition*, *37*(1), 1-19. https://doi.org/10.1016/j.patcog.2003.07.008