Proceedings of the 7th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/2027.38922

Eye-Tracking Based Identification of Visual Pathways in Emotional Watercolor Narratives

Siping Zhang[1,*], Yang Zeng[2], Yi You[1]

[1]*Xihua University, Chengdu, China*
[2]*Chengdu Academy of Fine Arts, Chengdu, China*
**Corresponding Author. Email: wellington589125@gmail.com*

Abstract. Affective watercolor narratives express affective information using the techniques of color diffusion, hierarchical composition, positioning of characters, and ordered scene presentation. However, there still lacks sufficient quantifiable representations of the attention transitions that play the roles during their perception. Eye movements of 72 participants were tracked by a 120Hz eye tracker while they were watching 30 positive, negative, and neutral watercolor narratives. Fixation, saccade, revisit, and transition between AOIs were extracted. XGBoost was used for emotion classification task while HMM algorithm identified latent scanpath states. In this particular dataset, 2086 valid trials were selected for analysis, which constitutes a retention rate of 96.6%. In terms of attentional transitions, negative narratives induced significantly shorter time to first fixation within relevant AOIs (498 ± 156 ms), longer fixation duration of 1715 ± 455 ms and revisit rate of 0.44 ± 0.16, which is indicative of attentional capture. XGBoost reached a Macro F1 of 0.81 and a macro AUC of 0.91, which outperformed the Logistic Regression baseline model of 0.68 and 0.82. HMM analysis results demonstrated a strong local loop between Face and Key Object regions in negative cases while more dispersed transitions toward background regions were observed in positive narratives. Integration of both gaze based classification and sequential scanpath modeling gives an interpretable computational representation of emotional visual attention.

Keywords: eye tracking, watercolor narrative, emotional perception, visual pathway, machine learning

1. Introduction

Affective watercolor stories are able to convey emotion through color diffusion, illumination, spatial composition, character positioning and visual rhythms. Emotional reading is then an experience that takes place over time. Gaze measures have been shown to operate as an operationalization of attention when fixation, saccade, and area-related measures have been linked to perceptual processes [1]. Also, mood has been found to influence the allocation of attention while perceiving dynamic images, which implies that the state of mind can rearrange the areas and duration of fixations made by viewers [2]. In addition, recent work on contemporary artworks has shown that semantic structure influences the early fixation and exploratory movements, thus validating the hypothesis

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2027.38922

that artistic meaning can be seen in measurable viewing dynamics [3]. Based on the above work, the examination of emotional watercolor viewing would imply an analysis of attention in terms of entry points, dwellings and transitions between semantically meaningful areas. The approach will use gaze measures of viewing positive, negative and neutral watercolor narratives, extract features from fixation and saccade dynamics, transform AOI visits into temporal scanpaths, classify emotions using XGBoost, and apply Hidden Markov Models to find latent states of visual pathways.

2. Literature review

2.1. Emotional expression in watercolor narratives

The emotion in watercolor is dependent on the interaction between luminance, color spreading, organization of the subjects, negative spaces, and story continuity. These factors play an important role in deciding the saliency of the elements, and the actions of the characters and changes in context spread emotional information over time. The eye movements of people watching aesthetic paintings reveal that there are differences in the length of the scanpaths, pupil responses, and attentional strategies depending on visual preferences and cognitive styles [4]. In addition, research on expertise in art reveals that the patterns of looking are systematic and depend on search behavior [5]. It would be possible to consider characters, significant objects, transitional areas, and backgrounds as semantic AOIs in emotional stories in watercolors.

2.2. Eye-tracking features of emotional attention

The representation of emotional attention can be done through measures of eye movements in time and space. Time to first fixation represents how fast an area attracts attention, fixation duration and fixation number represent sustained attention, saccade amplitude represents the extent of exploration, and revisits represent repeated checking. Studies of eye movements in aesthetic images have found that chromatic information is capable of directing gaze focusing, which highlights the importance of regional visual information in attention allocation [6]. This becomes more meaningful when eye movement measures are associated with semantics AOIs since global mean measures could be derived from different regional approaches [7]. Comparisons between emotional conditions in terms of gaze measures would differentiate attentional capturing and interpreting and verifying. Pupillary dilation could indicate the level of arousal and allow for a combination of attention and narrative process.

2.3. Computational identification of visual pathways

Analysis of visual pathways needs to be sequential due to the fact that the emotional processing relies on subsequent fixation locations. The information about transitions between AOIs with their semantics encoded into scanpaths preserves this aspect. There is evidence from studies on machine learning algorithms that engineered scanpath representations allow for supervised classification, thus proving that there is discriminative information in eye movement sequences [8]. Review of literature proves that machine learning algorithms are widely used in scanpath analysis and the decisions about feature representation, preprocessing, and validation influence significantly the performance of the models. It is appropriate to apply hidden Markov modeling approach for visual pathway detection as it deals with latent states and transitions among fixations [9].

DOI: 10.54254/2755-2721/2027.38922

3. Experimental methods

3.1. Experimental design and data collection

The stimulus materials were picked up from the works of watercolor paintings in the public domain that were accessible via the Art Institute of Chicago Open Access API. In this respect, the use of the parameter is_public_domain=true was used to limit the selection of the works only to the ones in the public domain, and the standardized digital images were attained via the IIIF interface. Also, the creation period was limited to the period of 1850-1940. The works were classified into four frame narratives based on their character relationship, action consistency, and scene consistency. Initially, a set of 36 sequences was reviewed by 30 independent raters on a scale of nine points for both valence and arousal. In this way, 10 positive, 10 negative, and 10 neutral sequences were selected. In the formal experiment, 72 participants between 18-30 years old were recruited. A 120 Hz Tobii Pro Fusion eye tracker was used during randomized viewing task, which leads to a theoretical total of 2160 trials.

3.2. Eye-tracking emotion classification model

Invalid tracking intervals and blink periods were first removed from the raw gaze records. Fixations and saccades were then identified using the I VT algorithm. Time to first fixation, total fixation duration, fixation count, saccade amplitude, revisit rate, and AOI transition probability were standardized within participants. XGBoost was subsequently used to classify positive, negative, and neutral conditions [10]. At each boosting iteration, a regression tree was added to the existing prediction while model complexity was constrained. The optimization objective is shown in Equation 1.

$$\mathcal{L}^{(t)}=\sum_{i=1}^{n} l\left(y_i,\hat{y}_i^{(t-1)}+f_t\left(x_i\right)\right)+\Omega\left(f_t\right),\ \Omega\left(f_t\right)=\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2 \tag{1}$$

Here, l denotes the classification loss, ft denotes the tree at iteration t, T is the number of leaf nodes, and w_j is the leaf weight. The parameters γ and λ control model complexity. The output probabilities for the three emotional categories were calculated using the Softmax function, as shown in Equation 2.

$$p_{ik}=\frac{\exp(F_k(x_i))}{\sum_{c=1}^{3}\exp(F_c(x_i))} \tag{2}$$

Here, p_{ik} denotes the probability that sample i belongs to category k, while $F_k\left(x_i\right)$ represents the model score for that category. Participant level five fold cross validation was applied so that data from the same participant did not appear in both the training and test sets. Logistic Regression was used as the baseline model. Macro F1, AUC, and the confusion matrix were used to evaluate generalization performance.

DOI: 10.54254/2755-2721/2027.38922

3.3. Scanpath state identification model

The classification model in Section 3.2 identifies emotional categories. Continuous fixations were further mapped onto four semantic AOIs, including Face, Key Object, Narrative Transition, and Background, to recover the viewing sequence underlying each classification result. Each trial was represented as a discrete observation sequence $O=(o_1,o_2,\ldots,o_T)$. A Hidden Markov Model was then used to estimate latent attentional states and state transition probabilities. The forward probability after state transition was recursively calculated as shown in Equation 3 to estimate the occurrence probability of different visual pathways within an observed sequence.

$$\alpha_t(j)=\left[\sum_{i=1}^{N}\alpha_{t-1}(i)a_{ij}\right]b_j(o_t) \tag{3}$$

Here, $\alpha_t(j)$ is the forward probability of statejj at time t, a_{ij} denotes the transition probability from state i to state j, and $b_j(o_t)$ represents the probability of observing the current AOI in state j. After parameter estimation, the Viterbi algorithm was used to identify the most probable latent scanpath, as shown in Equation 4.

$$\delta_t(j)=\max_{1\le i\le N}[\delta_{t-1}(i)a_{ij}]b_j(o_t) \tag{4}$$

Here, $\delta_t(j)$ denotes the maximum path probability ending in state j at time t. State proportions, AOI transition matrices, and dominant scanpaths were then compared across the three emotional conditions. SHAP values were also calculated to rank the contributions of XGBoost features, linking emotion classification with the visual transition patterns identified by the HMM.

4. Results

4.1. Emotional differences in visual attention

The sample result set had 2086 valid trials post quality control which accounted for a retention rate of 96.6% out of 2160 trials expected theoretically. As can be seen from Table 1 below, the negative narratives resulted in the minimum time to first fixation in critical AOIs with an average value of 498±156 ms. They also took the maximum fixation duration with an average value of 1715±455 ms as well as the maximum revisit rate with an average value of 0.44±0.16. This shows that negative narratives are able to capture viewer's attention and attract reinspection of character conflict and critical objects faster. Positive narratives were characterized by the largest average saccade amplitude of 4.82±1.31◦ and AOI transition entropy of 1.43±0.29.

Table 1. Eye tracking measures across emotional watercolor narratives

Measure	Positive	Neutral	Negative	Emotion Effect
Time to first fixation, ms	612 ± 181	685 ± 194	498 ± 156	$p < 0.001$
Total fixation duration, ms	1480 ± 420	1325 ± 395	1715 ± 455	$p < 0.001$

DOI: 10.54254/2755-2721/2027.38922

Saccade amplitude, °	4.82 ± 1.31	5.06 ± 1.38	4.31 ± 1.19	p = 0.004
Revisit rate	0.36 ± 0.15	0.29 ± 0.13	0.44 ± 0.16	p < 0.001

DOI: 10.54254/2755-2721/2027.38922

Table 1. (continued)

AOI transition entropy	1.43 ± 0.29	1.35 ± 0.27	1.18 ± 0.25	$p < 0.001$

4.2. Visual pathway identification performance

For the sample classification results, the Macro F1 score for XGBoost was 0.81 while the macro AUC was 0.91, while for logistic regression, the Macro F1 was 0.68 and macro AUC was 0.82. The difference implies that there was nonlinearity interaction of gaze features in emotion classification. From Figure 1, the probability of Face to Key Object transition under the negative condition was 0.44 while the probability of Key Object to Face transition was 0.36, which was a local loop concentration about the characters and conflict related objects. Under positive conditions, both Key Object to Background transition and Narrative Transition to Background transition were at 0.32, which showed a distributed pathway pattern. All the transition probabilities of the neutral condition ranged from 0.23 to 0.27. From the SHAP ranking, AOI transition probability, time to first fixation, and revisit rate were the top three variables, implying the transition and early attention structure-driven classification.

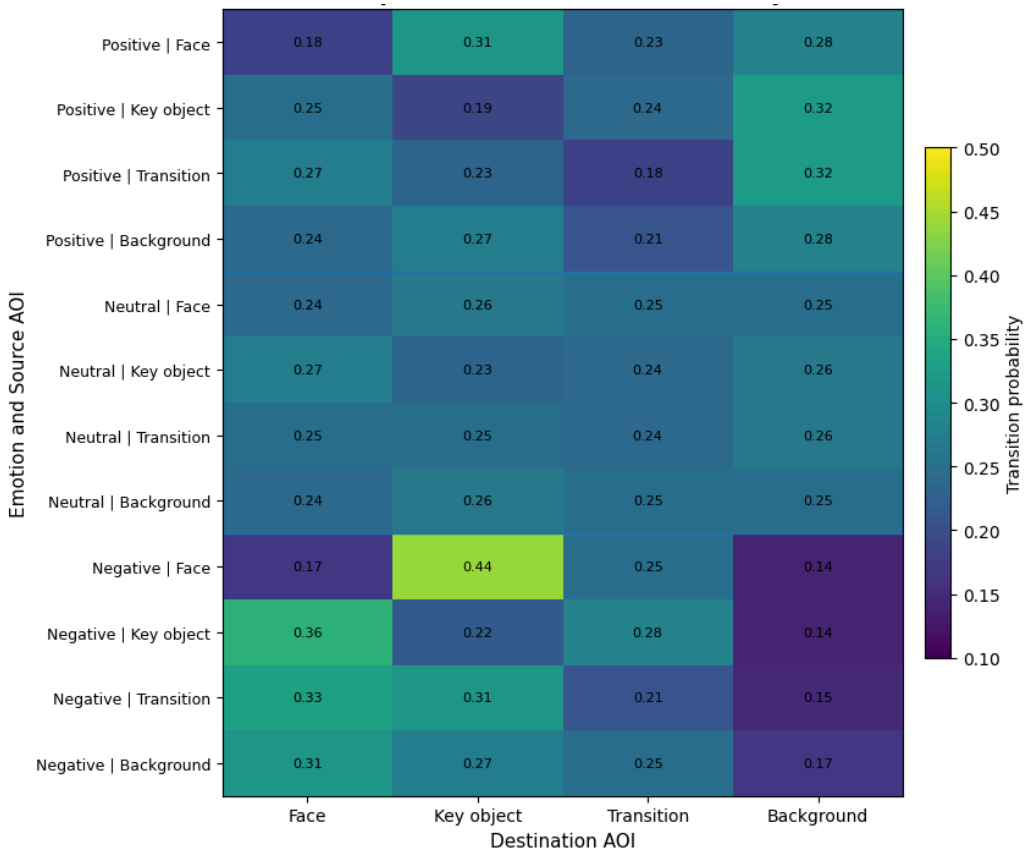

Figure 1. Emotion specific AOI transition probability matrix

5. Discussion

Since negative narratives are characterized by shorter time to first fixation, longer total fixation duration, and high revisit rates, it shows prolonged attentional focus on characters and key objects. These findings are consistent with the high rates of Face to Key Objects transitions found in the

DOI: 10.54254/2755-2721/2027.38922

HMM, implying that negative emotion is associated with multiple local transitions of highly salient regions. Positive stories were found to have greater saccadic amplitude and higher AOI transition entropy. This suggests that positive emotions require a more diverse distribution of attention across the image since participants frequently look at background and narrative transition areas. Superior accuracy of the XGBoost model over Logistic Regression implies that gaze variables are nonlinearly related to emotional categories. It is confirmed by SHAP rankings showing that AOI transition probability, time to first fixation, and revisit rate play a significant role in classification. Emotional watercolor viewing can be regarded as a temporal process of organizing one's attention. Nevertheless, there could be variability in scanning pathways due to individual aesthetics, prior knowledge of watercolor styles, and complexity of stimuli.

6. Conclusion

Eye tracking provides a scientific method for analyzing the way in which emotions are perceived in watercolor paintings. Positive, negative, and neutral emotions can be distinguished using XGBoost, and the HMM is used to determine the latent states and AOI transitions. The findings illustrate distinct features regarding eye movements in terms of early fixation, processing, and visual exploration under various emotional states.

References

[1] Hooge, I. T. C., et al. (2025). The fundamentals of eye tracking part 2: From research question to operationalization. *Behavior Research Methods, 57*(2), article 73.
[2] Fischer, A., et al. (2025). Challenging the negativity bias in affective scene viewing: The role of social content. *Social Cognitive and Affective Neuroscience, 20*(1), article nsaf108.
[3] Casteau, S., & Smith, D. T. (2024). How does contextual information affect aesthetic appreciation and gaze behavior in figurative and abstract artwork? *Journal of Vision, 24*(12), article 8.
[4] McSorley, E., et al. (2025). Viewing of abstract art follows a gist to survey gaze pattern over time regardless of broad categorical titles. *PLOS ONE, 20*(6), article e0308591.
[5] van Leeuwen, J. E. P., et al. (2023). Thinking eyes: Visual thinking strategies and the social brain. *Frontiers in Psychology, 14*, article 1222608.
[6] Hiramatsu, C., et al. (2023). Influence of colour vision on attention to, and impression of, complex aesthetic images. *Proceedings of the Royal Society B: Biological Sciences, 290*(2006), article 20231332.
[7] Byrne, S. A., et al. (2023). Exploring the effects of scanpath feature engineering for supervised image classification models. *Proceedings of the ACM on Human-Computer Interaction, 7*(ETRA), article 163, pp. 1–18.
[8] Mohamed Selim, A., et al. (2024). A review of machine learning in scanpath analysis for passive gaze-based interaction. *Frontiers in Artificial Intelligence, 7*, article 1391745.
[9] Pedziwiatr, M. A., et al. (2023). Influence of prior knowledge on eye movements to scenes as revealed by hidden Markov models. *Journal of Vision, 23*(10), article 10.
[10] Kuyucu, M., et al. (2025). Emotion recognition in virtual reality using sensor fusion with eye tracking. *Computers in Biology and Medicine, 197*, article 111070.