Article:
Kyle-Davidson, Cameron, Solis, Oscar, Robinson, Stephen et al. (2 more authors) (2025) Scene complexity and the detail trace of human long-term visual memory. Vision Research. 108525. ISSN 0042-6989

https://doi.org/10.1016/j.visres.2024.108525

Vision Research 227 (2025) 108525

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Scene complexity and the detail trace of human long-term visual memory

Cameron Kyle-Davidson *, Oscar Solis, Stephen Robinson, Ryan Tze Wang Tan, Karla K. Evans

University of York, Dept. of Psychology, York, YO10 5NA, UK

ARTICLE INFO

Keywords:
Long-term visual memory
Complexity
Neural networks
Scene perception
Scene memory
Scene detail

ABSTRACT

Humans can remember a vast amount of scene images; an ability often attributed to encoding only low-fidelity gist traces of a scene. Instead, studies show a surprising amount of detail is retained for each scene image allowing them to be distinguished from highly similar in-category distractors. The gist trace for images can be relatively easily captured through both computational and behavioural techniques, but capturing detail is much harder. While detail can be broadly estimated at the categorical level (e.g. man-made scenes more complex than natural), there is a lack of both ground-truth detail data at the sample level and a way to operationalise it for measurement purposes. Here through three different studies, we investigate whether the perceptual complexity of scenes can serve as a suitable analogue for the detail present in a scene, and hence whether we can use complexity to determine the relationship between scene detail and visual long term memory for scenes. First we examine this relationship directly using the VISCHEMA datasets, to determine whether the perceived complexity of a scene interacts with memorability, finding a significant positive correlation between complexity and memory, in contrast to the hypothesised U-shaped relation often proposed in the literature. In the second study we model complexity via artificial means, and find that even predicted measures of complexity still correlate with the overall ground-truth memorability of a scene, indicating that complexity and memorability cannot be easily disentangled. Finally, we investigate how cognitive load impacts the influence of scene complexity on image memorability. Together, findings indicate complexity and memorability do vary non-linearly, but generally it is limited to the extremes of the image complexity ranges. The effect of complexity on memory closely mirrors previous findings that detail enhances memory, and suggests that complexity is a suitable analogue for detail in visual long-term scene memory.

Introduction

Human visual long-term memory (VLTM) has a large storage capacity, with the capability to encode thousands of distinct images. If tasked to view 10000 images, a human, on average, could be expected to remember the majority of the images they were shown (Standing, 1973). Initially, this vast storage was thought to be due to the storing of gist traces: low-level, rapidly extracted information which carries global properties which allow for image categorisation. However, more recent work has shown that significant idiosyncratic detail survives the encoding process. Brady, Konkle, Alvarez and Oliva show that even when remembering large number of objects (Brady et al., 2008), performance in recognising the shown object remains high even when shown objects drawn from the same category. The ability to recognise a specific, previously seen object remains intact even with a large number of similar objects held in memory. Memory performance remains high with up to 16 same-category exemplars held in visual long-term memory (Konkle et al., 2010a). This performance is not exclusive to objects; in a later study Konkle et al. shows observers thousands of scene images, finding that memory performance remains high even with 64 same-category scenes held in memory. Moreover, this ability is robust: a doubling in exemplar amount results in only a 2% decrease in recognition accuracy. Despite having seen thousands of scenes, enough detail is preserved for each scene to allow an observer to determine they have previously seen a *specific* scene (i.e, a kitchen with these idiosyncrasies rather than another kitchen). This degree of recognition would be impossible if only a gist trace was stored. Generally, there is clear evidence that more than an impoverished gist trace of an object or scene persists in visual long term memory, (Brady et al., 2011; Cunningham et al., 2015; Guevara Pinto et al., 2020) instead, detail is encoded and preserved.

The gist component of human visual long-term memory has been the focus of much research (Larson et al., 2014; Oliva, 2005; Oliva & Torralba, 2006), and is reasonably well understood. In contrast, how the *detail* present in a scene interacts with that same scenes overall memorability is less known. Some scenes are filled with clutter, objects, and textural variations. Others might be comparatively plain:

* Corresponding author.
E-mail address: cameron.kyle-davidson@york.ac.uk (C. Kyle-Davidson).

https://doi.org/10.1016/j.visres.2024.108525
Received 5 March 2024; Received in revised form 30 October 2024; Accepted 21 November 2024
Available online 6 December 2024

consider the difference between a picture taken of a living room with that of an empty field - what impact does the presence or lack of detail in these respective scenes impact their memorability? Answering this question is difficult, largely due to the difficulty in extracting a reasonable representation of 'detail'. While the gist of a scene can be probed with rapid serial visual presentation, and extracted with straightforward computational statistical measures, no singular method exists for finding the detail component. Recently, a study by Evans and Baddeley (2018) proposes a two-level processing model for scene memory, employing visual complexity as an analogue for scene detail. The initial processing stage of the two-level model is based on gist, extracting general image features, whereas the second stage facilitates encoding of idiosyncratic scene elements. This work reveals that differences in image *detail* appear to affect how well a given image is remembered. Images of man-made scenes (indoor scenes etc, assumed highly detailed) are better remembered than natural scenes (outdoors, low detail). However, at the time of the study, there was not a ground-truth per-sample measure of detail for each individual scene, limiting further analysis of the relationship between detail and human memory.

Over the past decade the study of human visual memory has benefited from advances in the computer sciences. From detecting the dimensions associated with an image being remembered or forgotten (Bylinskii et al., 2015; Isola et al., 2011a, 2011b), to developing large scale neural networks for image memorability analysis (Khosla et al., 2015; Koch et al., 2020), to identifying and predicting the regions which cause a scene to be memorable (Akagunduz et al., 2019; Kyle-Davidson et al., 2019), computational techniques have aided psychological study into human memory. More recently, computational methods have been applied towards understanding of perceptual scene complexity, to understand which components of a scene image contribute towards a human believing that a viewed scene is either complex, or is simple (Kyle-Davidson et al., 2023). The level of complexity in a scene is driven by both low-level image properties as well as the semantic content of a scene; and hence is a highly tempting candidate to serve as an operationalisation of detail. It may be that the level of *complexity* in a scene, as perceived by a human has a direct interaction with the memorability of that scene, on a per-scene (rather than categorically grouped, such as man-made vs natural) basis, due to complexity capturing some degree of the detail trace of VLTM.

Complexity itself has a rich history, from original theoretical work (Birkhoff, 1933) to definitions based in line drawing detail (Snodgrass & Vanderwart, 1980) or verbal texture descriptions (Heaps & Handel, 1999). More modern approaches approximate the amount of information in an image with entropy calculations (Cardaci et al., 2009; Yu & Winkler, 2013), kolmogorov complexity (Kolmogorov, 1965; Rigau et al., 2007) or foveal clutter (Rosenholtz et al., 2007). State of the art approaches currently use machine learning models to generate complexity ratings for arbitrary images (Corchs et al., 2016; Kyle-Davidson et al., 2023; Nagle & Lavie, 2020). However, there have been few comparisons between complexity and memorability, with primary focus on comparisons of complexity and aesthetic measures (Sun et al., 2015; Van Geert & Wagemans, 2020, 2021), preference (Althuizen, 2021; Berlyne et al., 1968; Güçlütürk et al., 2016), or usefulness (Foster, 2010). Often, there appears to be an inverted U-shape relationship between complexity and these other perceptual characteristics. Certainly for aesthetics and preference metrics, low complexity and high complexity stimuli are less preferred, and less aesthetic, compared to their medium-complexity cousins. When it comes to memory, there is some evidence that an inverted U-shaped, non-linear relation also exists between stimulus memorability and stimulus complexity. Carlisle et al. find that images of medium complexity are best remembered in short term scene memory (Carlisle et al., 0000). Oliva et al. likewise suggest that long-term visual memory for scenes is not linear against complexity, and instead follows a similar inverted U, with medium complexity scenes being better remembered (Oliva, 2004). In contrast, a recent work comparing memory scores from a large dataset (though not scene focused) find that computational complexity measures correlate positively and solely linearly with hit rate (Saraee et al., 2020), with no evidence for non-linear behaviour. Nonetheless, there are suggestions of traceable neural correlates both for complexity (Güçlütürk et al., 2018), and for the impact of complexity on memory, with Chai et al. (2010) finding that complexity directly modulates activity in regions responsible for memory formation.

In this paper, we propose investigating the degree to which ground-truth complexity (drawn from human observers) interacts with ground-truth memorability data, and hence the impact of detail on how effectively a scene is remembered, or forgotten. We aim to narrow down whether complexity and memory vary together linearly, or non-linearly, over a broad scene dataset. In study 1, we evaluate the direct relationship between one and two-dimensional memorability and complexity data for scenes, and test the theory of two pathway encoding in visual long term memory. In study 2, we test whether a similar relationship also appears in an artificially predicted complexity measure, using a deep neural network to generate complexity ratings. Should predicted complexity ratings for an image align with the memorability of the image, this is indicative that complexity cannot be captured without also capturing the detail trace relevant to human memory. Finally, in study 3 we investigate the larger-scale relationship between complexity and memorability, giving rise to some general conclusions on how the level of detail present in a scene impacts the ability to remember that scene, and the mechanisms which appear to drive that memorability.

1. Study 1 - The relationship between complexity & memorability

To examine the relationship between complexity and memorability we draw upon two datasets which provide ground-truth memory, and ground-truth complexity data respectively for the same scene images. The recent development of this dataset has made the following analysis possible; prior to this, no datasets existed which provided both ground-truth memorability and complexity data for the exact same images.

The scene component of the dataset consist of 800 varied scene images across eight different categories (100 images per category). These categories span commonly encountered scenes, such as kitchens, living rooms, golf courses, amusement parks, and airport terminals, among others. Each image is 700 by 700 pixels, and preprocessed to minimise occurrences of overt text, recognisable landmarks, and faces looking directly at the camera. For both memorability and complexity data, both single-score overall data is provided, as well as two-dimensional annotations which indicate either regions relevant to the memorability of that scene, or regions relevant to the degree of complexity present in that scene. In this section we provide a brief overview of the methods used to obtain this data. Full details can be found in the relevant cited works.

Memorability data

The memorability dataset we use is VISCHEMA (Akagunduz et al., 2019). The VISCHEMA dataset consists of 800 scene images, evenly distributed among eight different categories. During the study phase of the experiment, participants ($n = 90$) were asked to remember 400 scene images randomly selected from the dataset. These 400 images were selected randomly, but constrained to ensure each image was seen as close to an equal number of times during the study as a whole. Each scene image was presented for three seconds. During the test phase, participants were shown another 400 images, of which 200 images were repeats from the study phase, and 200 images were foils that were not shown in the preceding phase. If the observer believes they have seen the displayed image before, they were instructed to press a button to indicate this. They were then asked to annotate up to three regions on the image that caused them to recall having seen that image previously. From this data, for each image an overall hit

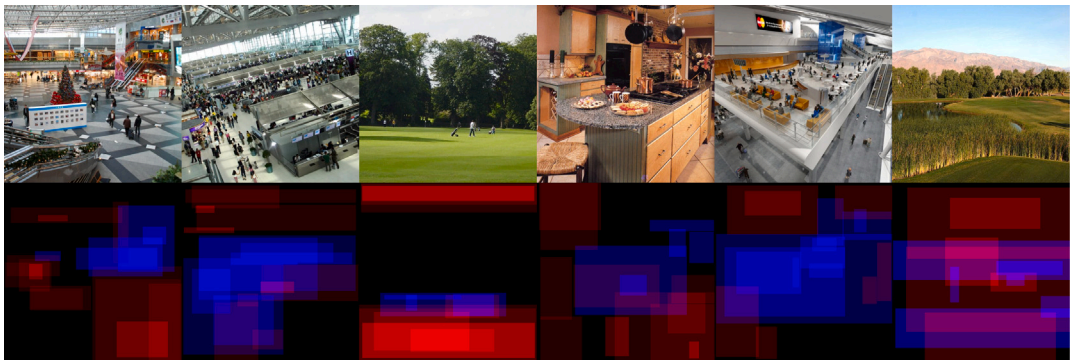

Fig. 1. VISC-C Complexity maps with complex regions in blue and simple regions in red.

rate, false alarm rate, and d-Prime was calculated. The annotation data reveals the two-dimensional regions of the scene that are indicated by an observer to be responsible for the scene being remembered (true memorability), or which cause an observer to believe they have seen that scene before, when in fact they have not (false memorability). In total, for the 800 scene images, this dataset comprises 800 single-score memorability ratings, and 800 two-dimensional maps which indicate the memorable and falsely memorable regions of the scene.

Complexity data

The complexity data we use in this analysis comes from the VISCHEMA-C dataset (Kyle-Davidson et al., 2023). This dataset contains single-score complexity ratings and two-dimensional complexity maps for the exact same scene images as used in the VISCHEMA dataset. Together, this provides a comprehensive corpus that allows for analysis of memory and perceptual complexity data. The complexity data was gathered via Prolific, an online experimentation platform (Prolific, 0000). Participants were shown a continuous stream of 200 randomly drawn scene images. The randomisation of each image stream was balanced to ensure the same number of participants viewed each image. The order of the stream was randomised to avoid context effects. Participants were first asked to rate the complexity of the image between 0 (very simple) and 100 (highly complex) on a sliding scale. Once they had rated the image numerically, they were then asked to annotate *either* the complex or the simple regions of the images, by being asked to draw free-form boxes around these regions. The same participant was not tasked with annotating both simple *and* complex regions for the same scene image. Participants showed a good degree of consistency between complexity ratings ($r = 0.84$). Complexity ratings also correlated well with the annotations given by the participants (multiple linear regression, $R^2 = 0.6$). This relationship was established by combining the two-dimensional annotation data from each participant which saw a given scene image into a 'complexity map' (Fig. 1) for that scene. These maps were then decomposed into two single-dimensional metrics. These were: (a.) Coverage: which describes how much of the image was covered by the simple or complex annotations from the participants (e.g., one might imagine a highly complex image to be complex throughout the scene rather than in one localised area) and (b.) intensity, which measures the degree of overlap between annotations (e.g., are the participants labelling the same areas as complex?), which is indicative of annotation consistency. These metrics show a high degree of correlation with the complexity score given by the participants, and shows that the annotations reliably capture the perceived areas of complexity or simplicity within that scene (Kyle-Davidson et al., 2023).

Results and discussion

We first compared the ground-truth complexity ratings with the corresponding ground-truth d-prime score for that image, and find a

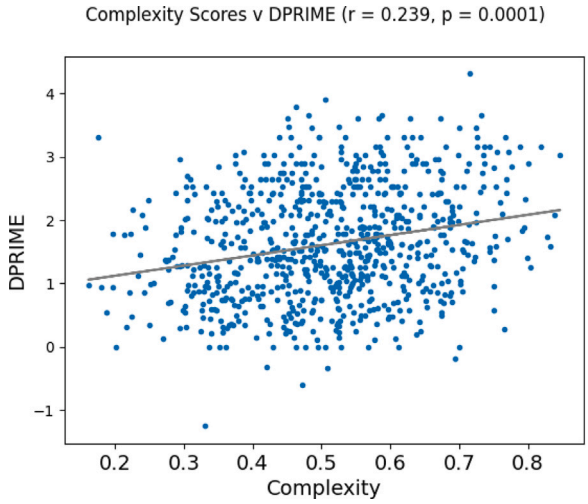

Fig. 2. Pearson's correlation between ground-truth human complexity ratings and ground-truth human memorability scores.

significant positive correlation (Fig. 2). That is, as the participants' complexity ratings increase, so too do the memorability ratings for those scenes.

However, d-prime alone does not necessarily tell the whole story; is the rise in memorability as complexity ratings increase carried by hit rate, false alarm rate, or both? To determine this we compare complexity ratings separately with both the hit-rate and false alarm rate for each image, shown in Fig. 3. Both the relationship between complexity ratings and hit rate, and complexity ratings and false alarm rate, is significant. A rise in complexity ratings corresponds with a rise in hit rate as well as with a decrease in false alarm ratings. This suggests that complexity (and correspondingly, detail), has a role in both increasing the likelihood of recognising an image, and decreasing incidences of false recognition. A greater level of detail appears to help prevent a viewed scene being confused with a previously encoded scene, due to the greater levels of idiosyncratic information that can be encoded, which corresponds with the data presented in studies of Evans and Baddeley (2018). This helps to separate the encoded image from the viewed image and helps reduce false recognition. The detail present may also facilitate correct recognition by providing more features that can be encoded and later recalled.

To gain further understanding of the relationship between image complexity and memorability we can investigate how the complex and simple image regions influence the hit rate and false alarm rate of image memorability. Is a scene more memorable because participants agree more on the complex regions of the image, or because more of the image itself is considered complex? Analysing the two-dimensional data

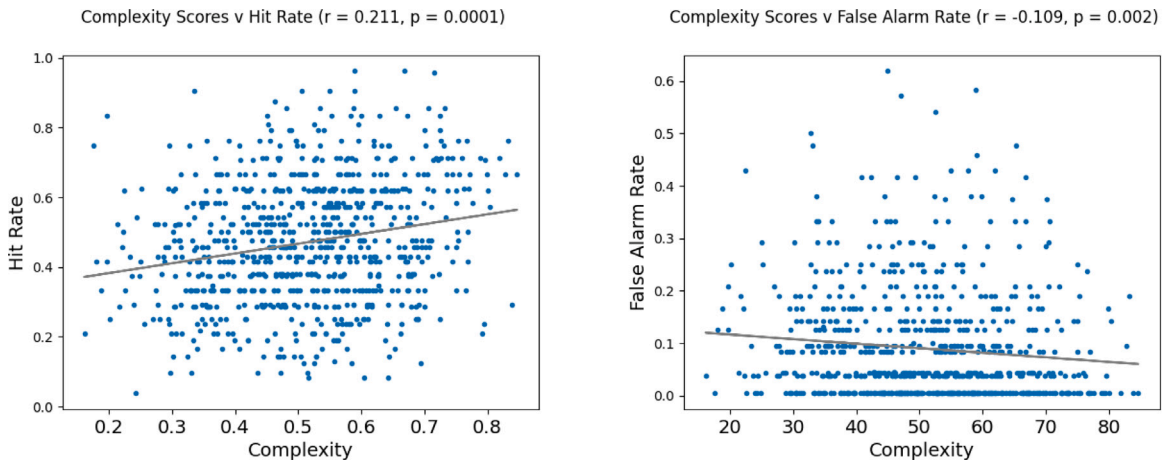

Fig. 3. Pearson's correlation between human complexity ratings and scene hit rates (left) and false alarm rates (right).

Table 1
Relationship between scene memorability (hit rate, false alarm rate) and metrics for describing two-dimensional scene complexity.

	Complex region intensity	Complex region coverage	Simple region intensity	Simple region coverage
Hit Rate	0.025	0.011	**−0.083***	−0.012
False Alarm Rate	**−0.089***	0.025	**0.124*****	**0.101****

Bold values are significant with * denoting $p < 0.05$, ** $p < 0.005$, and *** $p < 0.0005$.

Table 2
Results of multiple linear regression, with Complex and Simple coverage removed to avoid multicollinearity concerns. Coefficients for each variable are shown, as is the coefficient of multiple regression (R) and variance explained (R-squared), as well as the variance explained when including all factors (af-Adjusted). All regressions are significant. Complexity can explain a small, but significant portion of variance inherent in memorability data for DPrime, hit rate, and false alarm rate.

	D-Prime	Hit rate	False alarm rate
Constant	1.000	0.3156	0.098
Complex Intensity	−0.051	**−0.096***	−0.014
Simple Intensity	−0.360	0.01	**0.061***
Complexity Scores	**1.431*****	**0.355*****	−0.042
R	0.244	0.23	0.136
R-squared	0.06	0.053	0.018
Adjusted R-Squared	0.056	0.049	0.015
af-Adjusted R-Squared	0.068	0.06	0.027
Observations	800		

Significant values shown in bold, p < 0: ***, 0.05: *.

allows us to take a step towards determining the mechanisms by which complexity influences memorability by revealing how complexity and memorability interact spatially. Paired with two-dimensional memorability data, for example, we can ask whether a generally memorable scene region is also generally complex — or vice-versa. To investigate this we start by comparing (1.) the average intensity of the complex or simple regions, which is representative of observer consistency, and (2.) the percentage of the image covered in annotations, with the scenes memorability. We show these relationships in Table 1.

Complex region intensity (observer agreements in annotations, i.e. overlap on a specific region) does not show any significant relationship with the hit rate of image memorability (r = 0.025; p = 0.484) but does with the false alarm rate (r = −0.089; p = 0.011). This pattern indicates that the more agreement there is between observations on the annotations of the complex regions in the image, the fewer false alarms in image recognition. The relationship with the 2-D map intensity of 'simple' labelled regions shows a complimentary, but different pattern of results. We find a significantly negative relationship with the hit rate (r = −0.083, p = 0.019) and a positive relationship with false alarm rate (r = 0.124, p = 0.0001). Generally, the more agreed-on simple regions in the scene, the lower the hit-rate, and the greater the false-alarm rate. In contrast, the scene coverage of simple or complex annotations compared to hit and false alarm rate shows less consistent results and is less informative. The complex region coverage shows no significant relationship with either hit or false alarm rate. However, the simple region map coverage does show a significant positive relationship with false alarm rate (r = 0.101, p = 0.004), though no relation with hit rate (r = −0.012, p = 0.727).

All of these relationships can be better summarised if one conducts a series of multiple linear regression (MLR) analyses considering one of the memorability metrics (DPrime, hit rate, false alarm rate) with a series of complexity metrics (Complex/Simple channel intensity, and human complexity ratings). We show the results of the MLR in Table 2. We exclude the complex/simple coverage factors to reduce multicollinearity, as well as results that include all available factors ('af-Adjusted R-Squared'). We find that perceived complexity of an image can explain a small, yet significant, portion of variance in memorability scores such as d prime (5.6%); hit rate (4.9%) and false alarm rate (1.5%). Interestingly, the complexity measures explain much less of the variance in false alarms than in hit rates, likely a result of the greater degree of human variation in false alarms. The data from the multiple regression analysis supports the initial findings, that for memorability defined by d-prime, a significant predictor is the human complexity score ratings. Breaking this apart into hit rate and false alarm rate reveals that for hit rate, the primary predictor remains human complexity ratings, while for false alarms the most critical factor is the observers agreement on simple regions in the image. This reinforces our earlier finding that more simple scenes are more vulnerable to false alarms.

Finally, given that we also have two-dimensional image memorability data, we can also directly compare the two sets of maps (complexity & memorability). The memorability map data contains both a 'true schema' channel; indicating regions that caused the scene to be correctly remembered, and a 'false schema' channel, indicating regions that cause false remembering. The complexity data contains a 'complex' channel indicating complex regions, and a 'simple' channel indicating simple regions. The results are shown in Table 3, with all the Pearsons 2D correlation statistically significant. The strong positive correlation for regions labelled as perceptually complex with 'true schema' and 'false schema' regions suggests that the more perceptually complex a region is, the more likely we are to remember it — or believe we have seen it before. For perceptually simple regions the relationship with memorability (though statistically significant) is much weaker and negatively correlated. This suggests that the perceptually simpler the

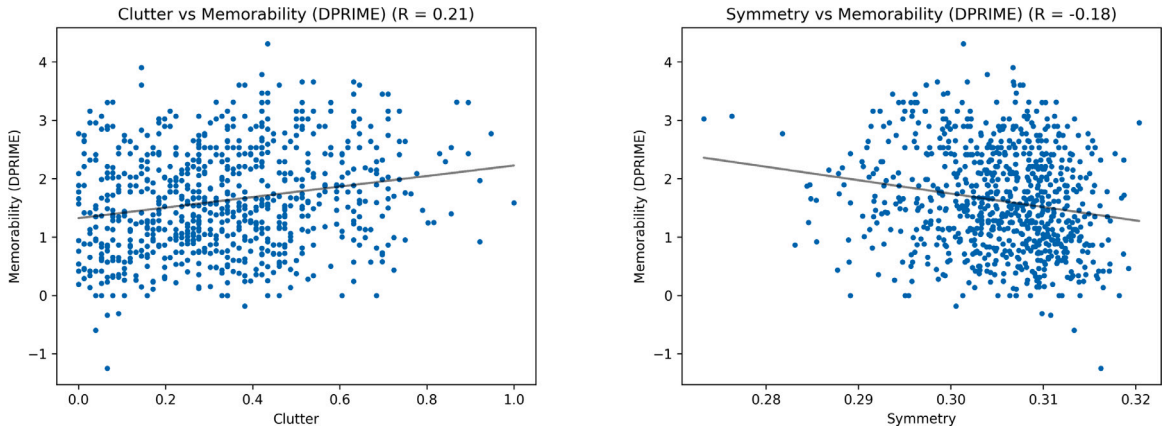

Fig. 4. Relationship between clutter (left) and symmetry (right) computational metrics and scene d-prime. All correlations significant.

Table 3
Comparing correlation of two-dimensional regions between memorability data and complexity data.

	Memorability		False memorability	
	Complexity	Simplicity	Complexity	Simplicity
ρ	0.5	−0.06	0.34	−0.05

All values significant.

region is perceived to be, the less likely it is for that region to be correctly or falsely be remembered as seen. To summarise; complex regions also tend to be memorable or falsely memorable, while simple regions are both less likely to be remembered and less likely to cause you to believe you have seen a scene when in fact, you have not.

Computational metrics

Prior work has shown that two computational images measures, based on psychological studies, perform reasonably well at explaining human perception of complexity. These measures are: (1.) visual clutter, computed via region-adjacency graph segmentation, which calculates the number of perceptually distinct regions in the scene, and (2.) patch symmetry, which computes how symmetric the image appears to be in a sliding window at varying scales. We compare how well these algorithmic complexity measures relate to the ground-truth memorability ratings of the VISCHEMA images by calculating the Pearson's correlation between these metrics (computed for each image) and the d-prime of each image. The results are shown in Fig. 4.

We find visual clutter, which is positively correlated with complexity, is also positively correlated with memorability as measured by d-prime. Symmetry, negatively correlated with perceived complexity, is likewise also negatively correlated with memorability. Generally, the more clutter and less symmetry present in a scene, the more memorable that scene, and vice versa.

Clearly there is a relationship between the perceived complexity of a scene image and that images memorability. This relationship appears both in direct comparisons of the ground truth data, is visible in exploratory modelling, and even shows up when comparing *computational* measures of scene complexity with memorability. Generally we find that the more complex the scene, the more memorable that same scene. This relationship is carried both in the hit rate and the false alarm rate of a given scene; the more complex the image, the greater the likelihood of a correct recognition, and the lower the chance of an incorrect recognition. A greater level of detail present in a scene may provide more potential features that can be encoded during the first time the image is viewed, which helps to both correctly identify a repeat of the scene, while also helping to filter out incorrect matches that lead to false recognition. Interestingly, prior work only finds that detail (considered between groups of manmade vs natural scenes) is only carried in false alarm rates (Evans & Baddeley, 2018). By considering complexity at an image level, rather than a category level, our approach allows for the capturing of complexity values that might be lost through grouping; for example, the existence of simple man-made images and complex outdoor scenes. This allows us to reveal the impact of detail on both false alarm rates and hit rates.

From the two dimensional data we find that when multiple participants indicate the same region as complex in a scene, that scene is also more likely to have reduced instances of incorrect recognition. For the simple channel, the data shows an inverse pattern; the more agreement on simple regions in the scene, the more likely that image is to be falsely recognised. This also carries a small, but significant reduction in hit-rate. Generally, it appears (when considered solo) that the simple regions in the image have more to do with the memorability of that image than the complex regions. However, this is not necessarily a complete picture. Examining the relationship between complexity-based annotations and memorability-based annotations, it is evident that the complexity of the region appears to drive the memorability of that same region. In essence, while simplicity appears to drive false-alarms up, when an image is correctly recognised, the regions that have caused this correct recognition are, in part, related to the complexity of that region. Even in a simple image, which should have a greater incidence of false alarms, the effect which causes correct recognition of that image is still partially to do with the complex regions present inside that scene. This makes sense, given the prior data on complexity; even simple images can contain potentially complex regions, a detail visible in the 2D data.

So far, this data suggests that complexity is a suitable operationalisation for the level of detail present in the scene. Previous work suggests that scene detail allows for recognition of an image amongst similar distractors, and leads to a reduction in false alarms. This is precisely the pattern we find for the interaction between complexity and memorability. This relationship is powerful enough to be carried in computational measures of complexity. We do not necessarily suggest that complexity *is* the entire detail trace; there could certainly be elements of 'detail' that are not captured in single-score and two-dimensional metrics of 'complexity' - however, complexity appears to be a suitable enough analogue to explain a small but significant portion of variance in memorability.

2. Study 2 - predicted complexity & memory

There is a clear relationship between the ground-truth memorability and ground-truth complexity data. To further verify this, we can explore whether a similar relationship arises in *predicted* complexity values. Essentially, we can ask whether the features that a machine learning model learns to use to predict the complexity of an

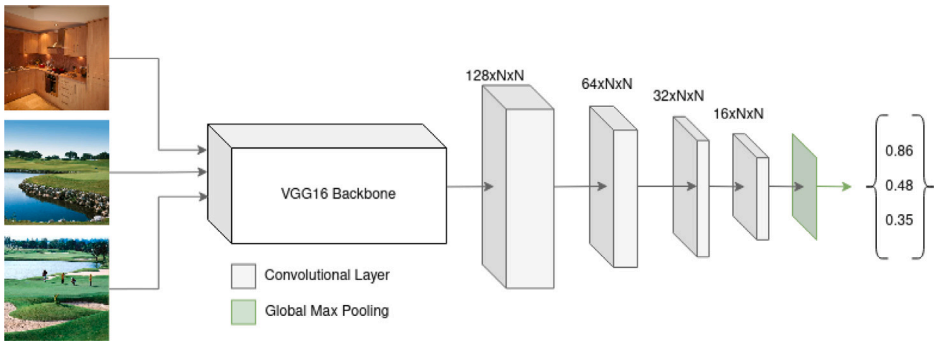

Fig. 5. Complexity prediction neural network derived from Kyle-Davidson et al. (2023).

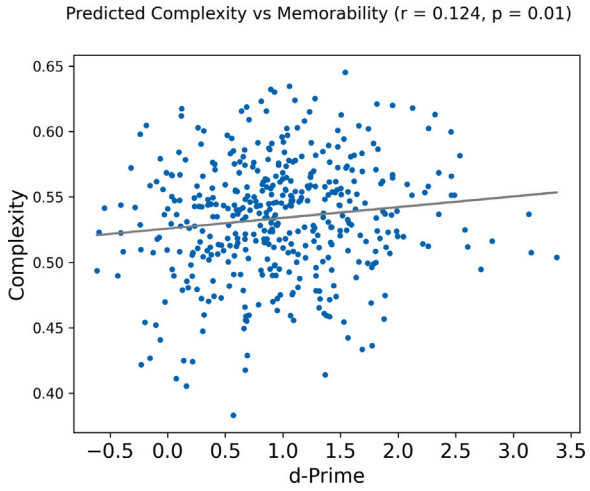

Fig. 6. Relation between ground-truth memorability data and neural-network predicted complexity values.

image also associate with the ground-truth memorability of the image. If so, model-predicted complexity ratings (for images that were not used to train the model) should show some degree of relation with memorability. This would imply that complexity and memorability are intrinsically intertwined — the model cannot learn to predict complexity without incorporating the relationship between complexity and memorability. If the model can learn this, it suggests that features which drive human complexity ratings may be intrinsically related to human memorability ratings. We are able to evaluate this due to recent development of neural network models for complexity prediction (Kyle-Davidson et al., 2022, 2023).

Deep neural network architecture

The neural network (Fig. 5) consists of a pre-trained VGG16 backbone for object detection, followed by four additional convolutional layers. The input to the network is a 224 × 224 × 3 scene image, and the output is a single score between 0 and 1.0 which indicates the level of complexity present in the scene. The network is fully convolutional, employing global max pooling in the last layer to reduce the feature dimensionality prior to the score output. The network uses ReLU activations throughout, aside from the final output which uses a sigmoid activation. In total the network is trained for 100 epochs with an RMSProp optimiser, with a learning rate of 0.0001. The loss function is straightforward mean-squared error between the predicted outputs and the ground-truth human data. To predict complexity scores eight-fold cross validation is used, each time predicting scores for an unseen portion of the data. The training was conducted with a single NVIDIA V100 GPU.

Results

The DNN has good prediction of complexity scores, with predicted scores achieving a Spearman's correlation of $\rho = 0.67$ with human complexity ratings. This allows for a reasonable comparison between predicted complexity scores and ground-truth memorability for those images (Fig. 6). It appears that computational measures of complexity, such as neural networks, sufficiently capture human complexity perception to also capture the relationship between complexity and memorability. That is, the relationship between complexity and memorability is robust enough that even predicted data shows a significant positive correlation ($r = 0.124$). This mirrors the ground-truth human data, and together suggests converging evidence that the relationship of the data as a whole is linear; that as the complexity of a scene increases, so does the overall memorability of that scene. This linearity apparent in both ground-truth and predicted data suggests that there are no deleterious effects of detail in scene images, and that the presence of increasing levels of detail only aids in the later recognition of that scene.

3. Study 3 - complexity & memory under load

The linearity of the complexity–memory relationship is surprising. While it makes sense that low-detail images are more poorly remembered due to lack of idiosyncratic detail, leading to easier confusion with similar distractors, one might expect that scenes that contain great levels of detail to also suffer a decrease in memory performance, due to the presence of a greater number of features, which could make encoding more difficult, and false alarming more likely. This would lead to the hypothesised inverted-U shape common to relationships between image characteristics. However, no evidence of this non-linearity is so far apparent. There are several potential reasons that so far the relationship has appeared linear. First, the complexity–memorability relationship may actually *be* linear. However, it seems unlikely that by increasing the detail in a scene image one can also increase that scenes memorability an eventual plateau. It is more likely this effect is due to one of two conditions: 1. Constraints on the dataset used, or 2. Noise inherent in human complexity/memorability data. Before making conclusive statements, these two conditions should aim to be ruled out.

While the dataset spans a broad collection of scene categories, it certainly does not contain examples of every *possible* scene. The multidimensional space defined by complexity and memorability scores is immense, and given that any scene image will lie somewhere in this space, it is highly unlikely that our dataset spans the entire range of possible complexity values for the entire possible space of scene images. We may be capturing a locally linear portion of a globally non-linear relationship. To determine if this is the case, we can artificially modulate the effect of complexity on human memory by introducing a factor of cognitive load into the repeat-recognition experiment. The increased load during the study phase of the experiment leads to difficulty encoding proportional to the level of load introduced.

This is effectively similar to introducing scene images to the dataset which are comparatively more difficult to remember, without having to actually obtain novel scenes, memorability, and complexity data. Note that we are not attempting to modulate *complexity* via load — but memorability. Prior work has shown that the additional difficulty imposed by load consequently means that the image features which contribute to or detract from a scenes memorability will have a greater impact on whether that scene will actually be remembered; that is, memory for images with lower levels of idiosyncratic detail will be disproportionately affected by load (Evans & Baddeley, 2018). Load hence broadens this 'local portion' of the complexity–memory space we are able to explore; capturing a wider range of the relationship. For the second condition, we can reduce the impact of noise on the complexity–memorability data by performing binning along the complexity axis — examining the average memorability of scenes that lay within a certain range of complexity (for example, between 0 and 15, 15 and 30, etc.). This will allow us to observe any larger dynamics within the data without these potentially small effects being drowned out by human noise.

Experimental design

Participants

The participants were undergraduate students from the University of York, recruited via opportunistic sampling, either by word-of-mouth, or via the University's Psychology Human Participant Pool System recruitment portal (SONA). 24 participants were recruited for each of the 3 conditions, for a total of 72 participants, based upon prior work which also investigated the effect of load on scene memorisation (Evans & Baddeley, 2018). We use Cohens F to perform an a-priori power analysis, finding a Critical F value of 3.98 and minimum sample size of 14. The mean age of participants was 20. Ten participants identified as male, while the rest identified as female.

Stimuli and apparatus

The experiment was coded using MATLAB R2021a and the Psychophysics Toolbox (Brainard & Vision, 1997) and ran on Microsoft Windows 10, using a Dell XPS computer. The images and instructions were presented on a Dell UltraScan P1110 CRT monitor with a face diameter of 21 inches, a resolution of 1280×1024, and a refresh rate of 85 Hz. Any auditory stimuli were played through a pair of headphones.

The visual stimuli were taken from the VISCHEMA image set, which is composed of 800 images of a range of scene categories. To ensure that images were used as targets and foils equally between participants, and reduce the chances of floor-performance due to fatigue, these images were randomly allocated to one of four equally-sized sets, each containing 200 images. Each participant was assigned a set of images as targets and a set of images as foils in a systematic way for counterbalancing. Images were presented at a size of 14.9° by 14.9° of visual angle.

The auditory stimuli was created by using voice recordings of research assistant SR reading out three digit numbers.

Procedure

Before the experiment began, participants were randomly assigned to the low-load, medium-load, or high-load conditions. Participants read an experiment information sheet, and were verbally briefed about the nature of the task by the experimenter. Participants were presented with another set of instructions on screen that explained the task in detail. In the study phase of the experiment participants had two task to complete. One was a verbal task of repeating or counting back out laud from a three digit number during the span of every 5 trails. The other one was visual to try and remember the images they were seeing each for 3 s.

The study phase of the experiment began with the presentation of an auditory stimulus: a male voice reading a three-digit number. Participants were required to complete a different task depending on the condition they were allocated to. These were: repeat the three-digit number out loud (low-load), count backwards from the number by one (medium-load), or count backwards from the number in threes (high-load) until the next number was presented. Experimenters kept track of participants' performance on the load task and noted any mistakes. While participants were completing their load task, they were simultaneously presented with a sequence of 5 images. Each image was presented for 3 s, before switching to the next image in the sequence. After the 5 images, a fixation cross was presented for 3 s, followed by the auditory presentation of the subsequent number in the sequence. This number sequence was randomly generated prior to the experiment. This process was repeated until all 200 images in their set were presented.

The test phase of the task was a two-alternative forced choice task in which participants were sequentially shown the 200 images presented during encoding, randomly intermixed with 200 unfamiliar images. They were instructed to press the 'a' and 'k' keys to indicate if the image was familiar or unfamiliar respectively. After the participant provided a response, the computer provided feedback regarding the accuracy of their answer. The next image in the sequence was presented upon a keypress. This process was repeated until all 400 images were shown and rated.

Analysis

To reduce the impact of noise and investigate large scale dynamics, the complexity data for each scene was automatically divided into evenly spaced bins across the entire range of possible complexity values. The memorability data was then averaged for each corresponding bin, giving the average memorability value for a range of scenes of similar complexity. In total, this results in 21 evenly spaced bins. When examining memorability behaviour at the extremes of complexity (low or high), we collect data from the lowest seven or highest seven bins, which comprises 33% of the bins. This range was selected to encompass the majority of the non-linear behaviour observed.

Linear and Polynomial regression models are fit to the data using ordinary least squares (OLS) for parameter selection, allowing for goodness of fit metrics to be compared between the models. To further confirm the results, we perform cross-validation, using the models for predictive purposes. We first divide each set of memorability data (binned by complexity) into a train and test set. The training set includes approximately 70% of the data of that bin, while the test set contains 30%. We then fit a set of Polynomial models from degree 2 (quadratic) to degree 11, and a linear model on the training set. We then predict the memorability values of the test set and compare to the ground-truth memorability data, using mean-squared error. This allows us to determine whether non-linear predictors offer better ground-truth predictive power than the equivalent linear predictor trained over the same data. We run the cross-validation 1000 times for each model and declare a non-linear predictor as superior should the mean-squared error of the predictor over the test set be lower than the linear predictor over the same data.

Results

Comparing human complexity data to human memorability data across all load-levels (Fig. 7) reveals an intriguing pattern. While the relationship appears mostly linear for scenes that display 'medium' levels of complexity, the relationship is distinctly non-linear for scenes that lie in the extremes of complexity; either highly simple, or highly complex. As scenes become very simple, or very complex, we see an *increase* in memorability that deviates from the 'expected' linear change. Highly complex scenes appear to allow for a boost in memorability for that scene; though this effect diminishes for scenes that are considered the most complex in the dataset. Likewise, very simple images appear to be able to be remembered better than their slightly more complex

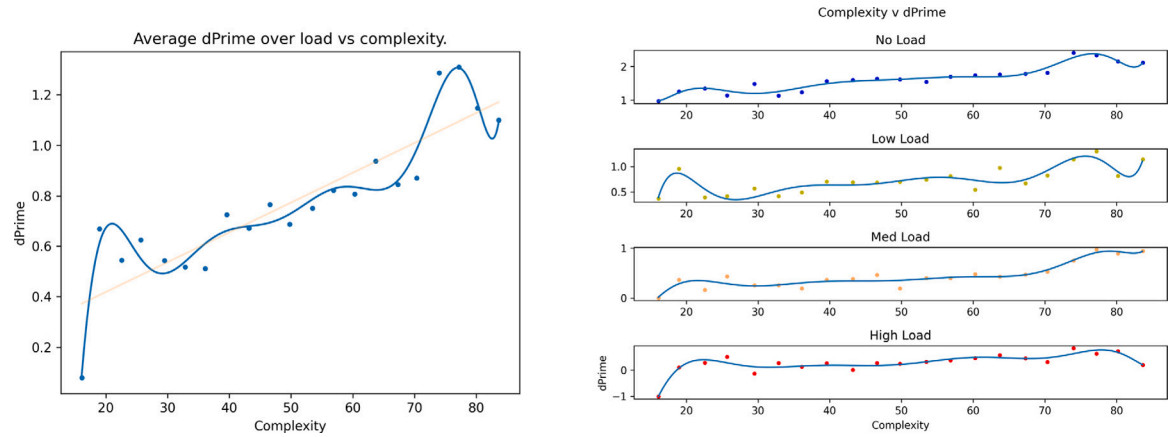

Fig. 7. Relationship between complexity and d-prime averaged over all loads (left) and broken down by load level (right). The data is fit with both a polynomial curve of degree 6 (blue), and a polynomial curve of degree 1 (a line, orange) for visualisation purposes. The data is follows a linear relationship except at the high and low complexity extremes; which deviate distinctly from simple linearity.

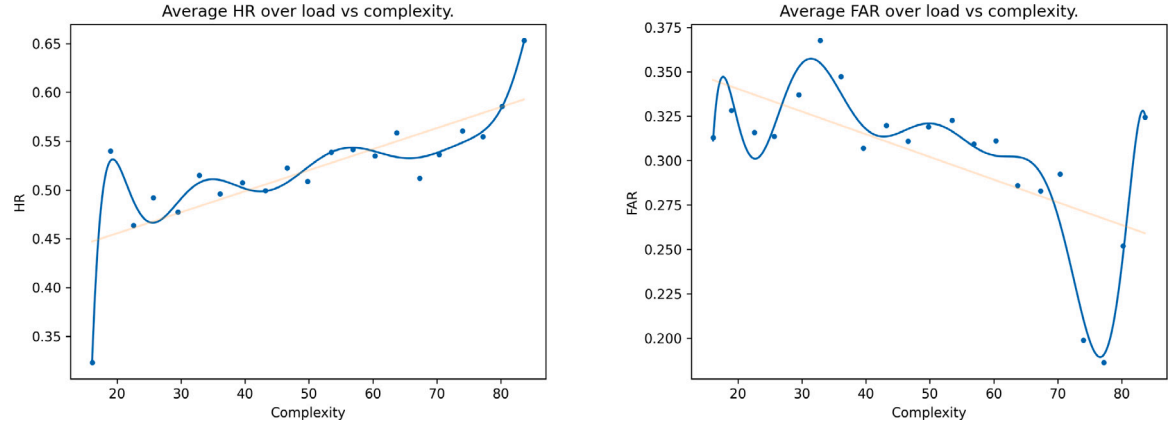

Fig. 8. Relationship between complexity and participant hit-rate (left) and false alarm (right) rate over all loads.

neighbours, with some indication that as scenes become extremely simple, memory performance falls. To determine statistically whether the data follows a linear or non-linear pattern we fit two models to the data, one linear model (polynomial of degree 1) and a set of non-linear models. Using these models we conduct an Ordinary Least Squares regression, and find that while a linear model captures the relation between complexity and memorability well ($R^2, 0.774$, likely due to the linear portion of the data) a polynomial model of degree 6 captures the relationship better ($R^2, 0.912$. This difference is more pronounced for hit rate (linear $R^2 = 0.563$, non-linear $R^2 = 0.815$) and even more obvious for false alarm rates (linear $R^2 = 0.387$, non-linear $R^2 = 0.855$).

Breaking this relationship down into hit rate and false alarm rate (Fig. 8) we see that while hit rate, for the most part, follows a generally linear trend, the picture for false alarm rate is more varied. Generally, scenes at the extremes of complexity, have false alarm rates that deviate strongly from the linear pattern shown by scenes that are neither highly simple nor highly complex. This is most obvious for the highly complex scenes, which show a significant (Kruskal–Wallis H-test, $H = 6.6$ $p < 0.02$) reduction (Mann–Whitney U, $U = 78.0$, $p < 0.01$) in false alarms compared to less complex images, though a similar effect is apparent for simplistic scenes. The presence of load does result in the characteristic decrease in memory performance, though we note that even for the high load conditions the observed D-Prime remains significantly different from chance (1-sample t-test, $p < 0.01$).

Examining the relationship across all load levels reveals a similar pattern occurring within each load (Fig. 7, right). Generally, scenes with low-medium to high-medium complexity follow a linear trend with regard to their memorability, irrespective of the level of load the participant was under. However, very low and high complexity scenes begin to deviate from this linearity. Generally, high complexity and low complexity scenes are remembered best, up until the level of complexity in the image crosses a certain threshold, and memorability decreases. This appears most prominently in the low and medium load conditions, showing an initial peak in memorability before a return to linearity, before becoming non-linear again as complexity increases. Interestingly, high load (3-back counting) appears to flatten this relationship back towards linear compared to lower levels of load. As the data from each load level is drawn from entirely separate groupings of participants, this suggests that the appearance of non-linear patterns in the complexity–memorability relationship is not a statistical fluke — and that complexity values in the extremes do indeed have a non-linear bearing on how well a given scene is remembered.

While the hit rates (Fig. 9, left) are relatively varied across the load levels, the false alarm rates (Fig. 9, right) show an interesting pattern: a reduction, then rise in false alarm rates for scenes with a complexity rating between 70–80. This gives rise to a characteristic 'dish shape' that appears in all load levels (with each load level having entirely distinct sets of participants). Here, reasonably high complexity scenes appear to cause a reduction in false alarm rates that deviates from the standard linear decrement observed for the majority of the data. Once the complexity increases further, the false alarm rate increases again. This appears to suggest that a high level of detail, to a certain extent, protects against false alarming. The effect of complexity on false alarms appears to be the primary source of non-linearity in the relation between memorability and scene detail.

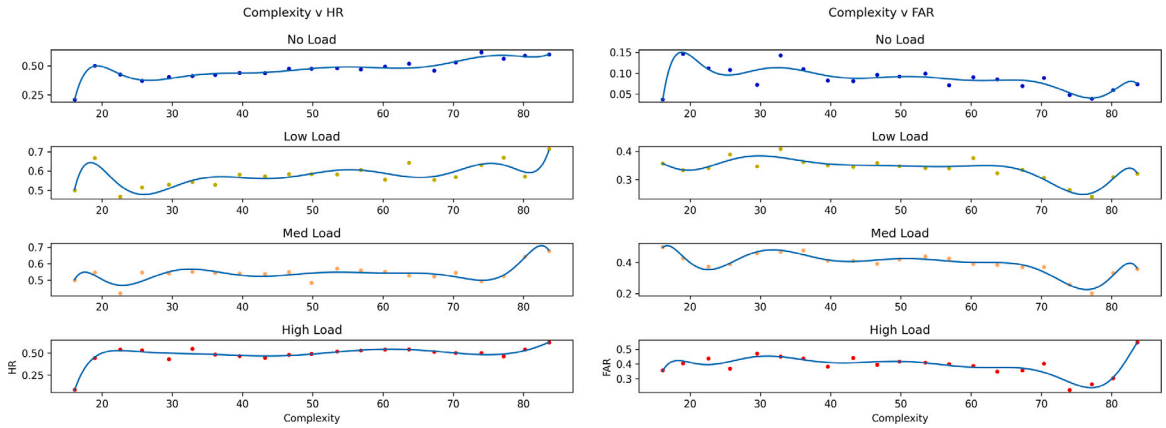

Fig. 9. Hit rate (left), false alarm rate (right), and complexity for all loads.

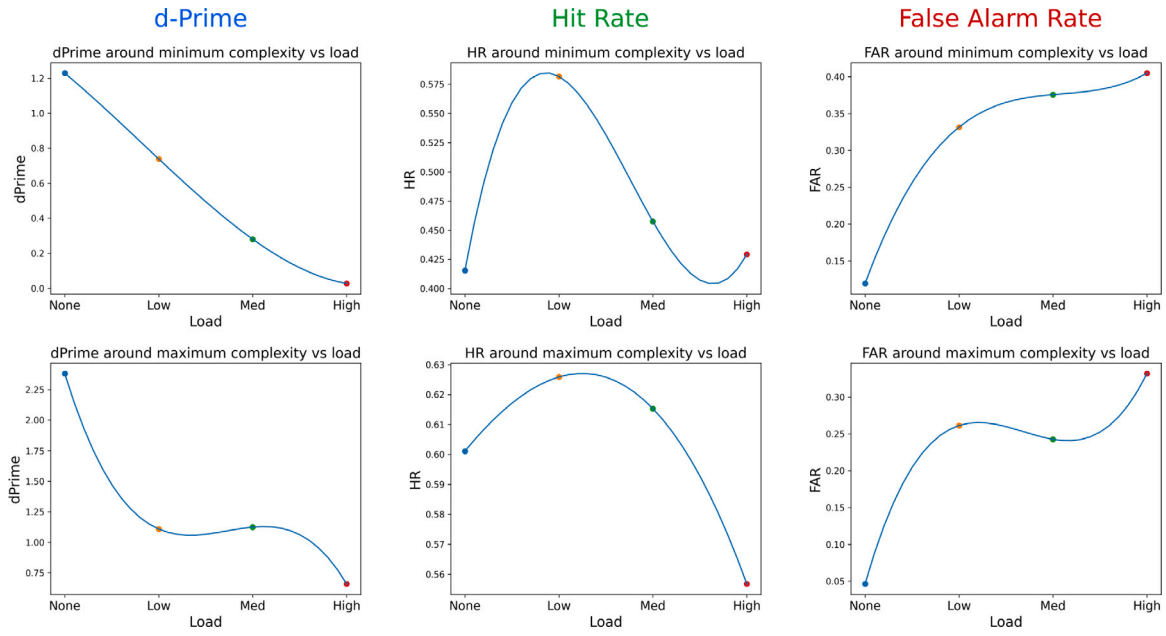

Fig. 10. Graphing complexity–memorability behaviour at the extremes of the complexity spectrum over load. We show d-prime behaviour (left column), hit rate behaviour (middle column) and false alarm rate behaviour (right column). The top row of graphs are for low-complexity images, while the bottom row are for high-complexity images.

So far these correlations are exploratory. To check how reliable the non-linearity of the complexity/memorability relationship is, we can instead perform cross-validation, training both linear and non-linear predictors on random subsets of the data. If the data truly does follow a non-linear pattern, then a non-linear predictor will have a greater accuracy then a linear predictor for the same subset of test data (data not used to fit either model). Comparing a linear predictor to a non-linear predictor for 1000 random subsets of the data, using polynomial models with degrees from 2 to 11, we find that there is always a non-linear predictor with superior performance to a linear predictor over the same training data. This is indicative that the relationship is best captured via techniques that account for non-linearity.

Behaviour at extremes

To explore this further we can examine the behaviour of the complexity–memorability relationship only in the regions that display empirical evidence of non-linearity. In Fig. 10 we graph load level, against the average d-prime, hit rate and false alarm rate of scenes that have either an unusually high or unusually low complexity, considering approximately 100 scenes for both the high and low complexity conditions. For scenes with low complexity, as load increases, there is an obviously linear decrease in memory performance, almost hitting floor performance under high load. The more cognitive load present, the worse you remember simple scenes. However, for highly complex scenes, performance does *not* decrease linearly with load. While there is an expected cost in memory performance when going from no load to low load, further increasing the load does not immediately lead to poorer memory performance for complex scenes. Instead, there is a plateau in performance, only falling again when the load level increases to high.

Breaking this down again into hit rate and false alarm rates, we observe a similar pattern for both complex and simple scenes in the hit rates. As load increases, hit rate increases, before plummeting as the degree of load rises, likely due to the participants becoming more liberal in their judgments. More interesting is the false alarm behaviour. Here, for low complexity scenes as load increases, false alarm rate increases. While this increase does not appear totally linear, it is monotonic: the more load, the more likely you are to believe you have seen a simple scene before, when in fact you have not. On the other hand, for highly complex scenes, there is again a plateau in false alarming between the low and medium load conditions. Generally, replicating the data above, the differences in memorability between

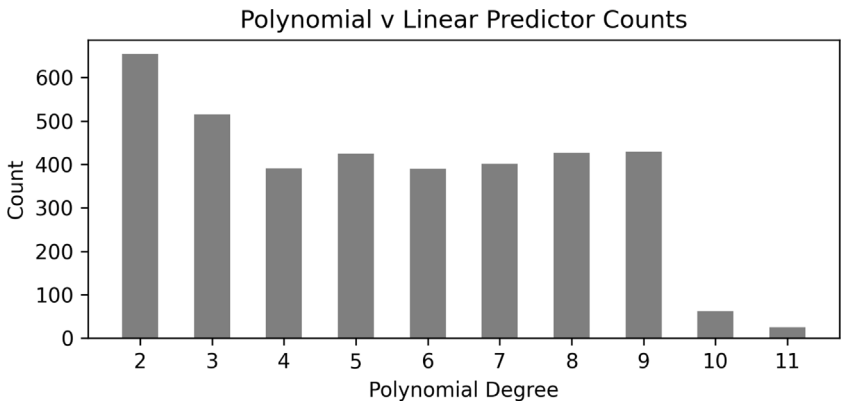

Fig. 11. Times out of 1000 that a non-linear predictor of degree *n* outperformed a linear predictor over the same data. Note that this is not exclusive — the linear model may have been outperformed by one, many, or none of the non-linear models.

high and low complexity scenes appear to be primarily carried by the false alarm rates. That is, you are less likely to remember having seen a complex scene under loaded conditions, when in fact you never saw that scene to begin with.

Cross-validation

We run a thousand-fold cross validation to determine the efficacy of non-linear predictors vs linear predictors over random subsets of the data. Generally, if the test data subset is best predicted by a linear model, this serves as evidence that the relationship tends towards linearity. Alternatively, if a non-linear performs better: the data may tend towards non-linearity. While this test is not as robust as identifying best-fit models on the entire dataset, it allows for further sanity checking. We find that out of 1000 random train/test splits, in approximately 80% of these splits a non-linear predictor outperforms a linear predictor trained on the same training data, and over the same test split. A breakdown of superior non-linear predictors is shown in Fig. 11. Generally, the performance of a quadratic model tends to beat linear, despite being the simplest non-linear model available. However, even if we exclude quadratic models, a linear predictor is still out-performed 69.8% of the time by an alternative non-linear model.

Discussion

Despite extensive research into both complexity perception and visual long term memory, there is surprisingly little work exploring the interaction between complexity and scene memory. There is somewhat more literature which examines the impact of *detail* on memory. Generally, this has focused on the detail retained for objects present in natural scene images (Hollingworth, 2005; Hollingworth & Henderson, 2002). However, more recently this has extended to examining the level of detail retained for the scene image as a whole (Konkle et al., 2010b), suggesting that the detail in the scene has an important role in preventing a remembered scene from being confused by similar distractors. However, 'detail' is a difficult metric to quantify. Here, we suggest that scene complexity serves as a suitable analogue for the more abstract 'detail' content present in an image, either capturing the detail trace (even in part), or varying in concert with it. Complexity is a natural tool for this task: it is difficult to imagine a natural scene which is undetailed, yet complex — or simple, yet containing a vast amount of detail. Recent work has made progress in both operationalizing complexity, and furthering understanding of the perception of it (Kyle-Davidson et al., 2023). However, when it comes to complexity and memory, investigations have yielded somewhat conflicting information: some evidence suggests medium complexity images are better remembered (Carlisle et al., 0000; Oliva, 2004) compared to low or high complexity images, whereas others suggest the relationship to be strictly linear (Saraee et al., 2020). This may be partially due to lack of data, smaller sample sizes, or due to complexity values being simulated.

In this work we attempt to address these shortcomings, analysing the relationship between complexity and memory for a scene dataset with a wide variety of categories and images, containing both ground-truth complexity and ground-truth human memory data. Given the interaction of complexity and other image characteristics, and earlier work on complexity and memory, we hypothesise that we should find a U-shaped relationship between complexity and memory for scenes. That is, simple scenes may be easily forgotten, and extremely complex scenes may fail to be encoded robustly. Instead, scenes of medium complexity may be best remembered; having enough detail to be easily separable from distractors, but neither too little idiosyncratic detail, or too much. We attack this hypothesis via three different studies: first examining the direct relationship between complexity and memory, then via artificial neural networks, and finally by using load to stress the impact of complexity on memory.

Perhaps most surprising is that despite prior hypotheses suggesting that complexity and memorability should have some form of U-shaped relationship (with medium complexity scenes remembered best), we find that the relationship to be strongly linear. The more complex (i.e the more detail), the better the scene is remembered. This holds both for the single-score ratings given by human observers, as well as in the two-dimensional maps attached to the scene images; regions that are complex often have a significant overlap with regions that are memorable. This suggests both that detail in a scene is advantageous, and that even very high levels of detail are not detrimental. High complexity in a scene may simply offer more idiosyncratic details to be encoded, allowing the cognitive representation of that scene to contain multiple distinctive features, which afford easy recognition of that scene. This aligns well with both (Evans & Baddeley, 2018; Konkle et al., 2010b), who suggest that high-fidelity encoding result in minimal levels of interference. As high levels of complexity would support high-fidelity encoding, it makes sense that higher complexity images are better remembered; an image of a field may be easily confused with other images of fields, or even forgotten entirely. A detailed image of a living room is less likely to be confused with other living rooms. This divide between the memorability of man-made and natural scenes is well known (Evans & Baddeley, 2018), and it is unlikely to be coincidence that simple scenes tend to be natural, and complex scenes man-made (Kyle-Davidson et al., 2023). It hence appears that during the encoding process, the available idiosyncratic detail that *can* be encoded is relatively greater for highly complex scenes vs average complexity scenes. This leads to a greater likelihood of a higher-fidelity memory trace for that more complex scene. Later, during retrieval, the

presence of the improved trace enables you to both identify that the re-occurring scene is a repeat (rise in hit rates), while simultaneously insulating you from the effect of similar distractors (fall in false-alarm rates).

The relationship between complexity and memorability is robust enough to appear even in simulated data. A neural network trained to predict complexity for scene images, even over scenes it was never trained on, still shows a linear positive relationship between predicted complexity scores and ground-truth memorability scores for those images. This implies that the same scene features a neural network might use to predict complexity are in some fashion related to the memorability of that scene. That is, memorability and complexity are not easily disentangled: by learning to predict one, you also learn (to a lesser degree) to predict the other. Finding the same pattern is encouraging - for a large-scale scene dataset, it seems highly likely that the memorability of a scene is related to its memorability. However, like the ground-truth human data, there is little evidence of any form of non-linearity, and no inverted U shape suggesting medium-complexity scenes are the most memorable.

It is possible that the data we are obtaining is a result of the scene dataset being used. While the entire space of scenes may contain a non-linear relationship between memorability and complexity, we are using only capable of using a small portion of that space. It may be that in this portion the relationship is broadly linear; only becoming non-linear at complexity values outside of the ranges available in our scenes. To explore this further, we examine the effect of cognitive load on the complexity–memorability relationship. Load during the memorisation phase increases the impact of detail, and effectively increases the range of complexity values available. We also employed histogram binning to gain a low noise overview of the data. We find that the relationship between complexity and memorability is indeed highly linear (even under high load) except for two critical sample sets: scenes of unusually high or low complexity. Scenes with complexity values that lay closer to the mean (complexity is Gaussian distributed, (Kyle-Davidson et al., 2023)) generally follow a linearly increasing trend. However, there are significant deviations from this at the extremes. Here, low complexity and high complexity scenes appear *more memorable* than might be expected from a linear relationship. This is especially obvious in the high-complexity cases. The effect appears to be majority driven by the impact of complexity on false alarms. Generally, increased complexity (i.e detail) reduces the likelihood of false alarming on a scene. At a certain high level of detail, this likelihood plummets, before rising again as the scenes become yet more detailed; moving beyond this 'sweet spot'. If we examine the extremes specifically over load, we find that while d-prime for low complexity scenes decreases linearly as load increases, for highly detailed images we observe a 'plateauing' effect, where additional load *does not* result in a decrease in d-prime. Again, this is carried in the false alarms.

Evans and Baddeley (2018) suggest that the encoding process of visual long-term memory consists of two separate processes. One of these processes occurs very rapidly, extracting a general description of a scene, whereas the other trace is slower, but extracts idiosyncratic detail. This trace is affected by both intention to remember and vulnerable to executive load, and is responsible for detecting a target from similar distractors. Critically, this 'detail trace' allows for robustness against false positives; effects which manipulate this detail process appear in the false alarms, rather than in the detections. Encouragingly, this aligns closely with the data we find when considering the relationship between memorability and complexity: high complexity images are more robust against false alarms than their low complexity counterparts. Indeed, most effects of complexity appear in the false alarm data rather than in the hit rate data. This is precisely what would be expected if complexity captured the detail trace. Furthermore, where Evans and Baddeley (2018) finds that memory performance for less detailed images is dependent upon availability of executive capacity, here we also show that as load increases, memory performance falls sharply for simple scenes, yet stabilises for scenes with high degrees of complexity. Effectively, high complexity in a scene image appears to provide a protective effect against false alarming, even under loaded conditions. Whereas in Evans and Baddeley (2018) these differences were shown on a categorical level, here we find a similar pattern for data with ground-truth complexity scores for each scene.

In conclusion, given our results, there is significant evidence that perceptual complexity can serve as a suitable analogue for the detail trace in visual memory. While we find little evidence that scene complexity and scene memorability follow an inverted U shaped curve, we do find that the relationship is non-linear, but more complex than prior literature might suggest. Scenes with high levels of complexity and remembered better than the linear trend would suggest, whereas scenes with *very* high levels of complexity fall back to the linear trend. Likewise, scenes with a minimal (but above floor) level of complexity are more easily remembered than scenes slightly more complex. In short: for the majority of scenes, as complexity increases, so does the overall memorability of that scene, likely due to enhanced interaction with the visual long-term memory 'detail process'. It is this pattern which drives the initially found linearity in the data. However, there are deviations from this towards the extremes: highly complex and highly simple scenes. We find that complexity itself protects against false alarming, aligning closely with the findings of both Evans and Baddeley (2018), Konkle et al. (2010b). Our data both indicates that detail can be successfully operationalised at the sample level, rather than the categorical level, allowing for future fine-grained analysis, and also supports the two-stage processing model for visual long-term memory.

Ethical approval

All studies performed as a consequence of this work were approved by the ethics board of the University of York, UK.

CRediT authorship contribution statement

Cameron Kyle-Davidson: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis. **Oscar Solis:** Writing – review & editing, Writing – original draft, Data curation. **Stephen Robinson:** Writing – review & editing, Writing – original draft, Data curation. **Ryan Tze Wang Tan:** Writing – review & editing, Writing – original draft, Data curation. **Karla K. Evans:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was funded by CRUK and Engineering & Physical Sciences Research Council (EPSRC), UK award EDDCPJT/100027 to Karla K. Evans

Data availability

Data will be made available on request.

References

Akagunduz, E., Bors, A. G., & Evans, K. K. (2019). Defining image memorability using the visual memory schema. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(9), 2165–2178.

Althuizen, N. (2021). Revisiting Berlyne's inverted U-shape relationship between complexity and liking: The role of effort, arousal, and status in the appreciation of product design aesthetics. *Psychology & Marketing*, *38*(3), 481–503.

Berlyne, D. E., Ogilvie, J. C., & Parham, L. (1968). The dimensionality of visual complexity, interestingness, and pleasingness. *Canadian Journal of Psychology/revue Canadienne De Psychologie*, *22*(5), 376.

Birkhoff, G. D. (1933). *Aesthetic measure*. Harvard University Press.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 4.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329. http://dx.doi.org/10.1073/pnas.0803390105.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*, *116*, 165–178.

Cardaci, M., Di Gesù, V., Petrou, M., & Tabacchi, M. E. (2009). A fuzzy approach to the evaluation of image complexity. *Fuzzy Sets and Systems*, *160*(10), 1474–1484.

Carlisle, N. B., Mack, M. L., & Oliva, A. The role of complexity in short-term scene memory. Citeseer.

Chai, X. J., Ofen, N., Jacobs, L. F., & Gabrieli, J. D. (2010). Scene complexity: influence on perception, memory, and development in the medial temporal lobe. *Frontiers in Human Neuroscience*, *4*, 1021.

Corchs, S. E., Ciocca, G., Bricolo, E., & Gasparini, F. (2016). Predicting complexity perception of real world images. *PLoS One*, *11*(6), Article e0157986.

Cunningham, C. A., Yassa, M. A., & Egeth, H. E. (2015). Massive memory revisited: Limitations on storage capacity for object details in visual long-term memory. *Learning & Memory*, *22*(11), 563–566. http://dx.doi.org/10.1101/lm.039404.115.

Evans, K. K., & Baddeley, A. (2018). Intention, attention and long-term memory for visual scenes: It all depends on the scenes. *Cognition*, *180*, 24–37. http://dx.doi.org/10.1016/j.cognition.2018.06.022.

Foster, E. M. (2010). *The U-shaped relationship between complexity and usefulness: a commentary*. American Psychological Association.

Güçlütürk, Y., Güçlü, U., van Gerven, M., & van Lier, R. (2018). Representations of naturalistic stimulus complexity in early and associative visual and auditory cortices. *Scientific Reports*, *8*(1), 3439.

Güçlütürk, Y., Jacobs, R. H., & Lier, R. v. (2016). Liking versus complexity: Decomposing the inverted U-curve. *Frontiers in Human Neuroscience*, *10*, 112.

Guevara Pinto, J. D., Papesh, M. H., & Hout, M. C. (2020). The detail is in the difficulty: Challenging search facilitates rich incidental object encoding. *Memory & Cognition*, *48*, 1214–1233.

Heaps, C., & Handel, S. (1999). Similarity and features of natural textures. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 299.

Hollingworth, A. (2005). The relationship between online visual representation of a scene and long-term scene memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(3), 396.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 113.

Isola, P., Parikh, D., Torralba, A., & Oliva, A. (2011). *Understanding the intrinsic memorability of images: Tech. rep.*, MASSACHUSETTS INST OF TECH CAMBRIDGE.

Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? In *CVPR 2011* (pp. 145–152). IEEE.

Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015). Understanding and predicting image memorability at a large scale. In *2015 IEEE international conference on computer vision* (pp. 2390–2398). IEEE, http://dx.doi.org/10.1109/ICCV.2015.275.

Koch, G. E., Akpan, E., & Coutanche, M. N. (2020). Image memorability is predicted by discriminability and similarity in different stages of a convolutional neural network. *Learning & Memory*, *27*(12), 503–509.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition ofinformation'. *Problems of Information Transmission*, *1*(1), 1–7.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, *139*(3), 558.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*(11), 1551–1556.

Kyle-Davidson, C., Bors, A., & Evans, K. (2019). Predicting visual memory schemas with variational autoencoders. arXiv preprint arXiv:1907.08514.

Kyle-Davidson, C., Bors, A. G., & Evans, K. K. (2022). Predicting human perception of scene complexity. In *2022 IEEE international conference on image processing* (pp. 1281–1285). IEEE.

Kyle-Davidson, C., Zhou, E. Y., Walther, D., Bors, A. G., & Evans, K. (2023). Characterizing and dissecting human perception of scene complexity. *Cognition*, *231*, Article 105319.

Larson, A. M., Freeman, T. E., Ringer, R. V., & Loschky, L. C. (2014). The spatiotemporal dynamics of scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 471.

Nagle, F., & Lavie, N. (2020). Predicting human complexity perception of real-world scenes. *Royal Society Open Science*, *7*(5), Article 191487.

Oliva, A. (2004). Complex scene images are simple in memory. *Journal of Vision*, *4*(8), 877.

Oliva, A. (2005). Gist of the scene. In *Neurobiology of attention* (pp. 251–256). Elsevier.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36.

Prolific: https://www.prolific.co/.

Rigau, J., Feixas, M., & Sbert, M. (2007). Conceptualizing Birkhoff's aesthetic measure using Shannon entropy and Kolmogorov complexity. In *Computational aesthetics* (pp. 105–112).

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*(2), 17.

Saraee, E., Jalal, M., & Betke, M. (2020). Visual complexity analysis using deep intermediate-layer features. *Computer Vision and Image Understanding*, *195*, Article 102949.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174.

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207–222. http://dx.doi.org/10.1080/14640747308400340.

Sun, L., Yamasaki, T., & Aizawa, K. (2015). Relationship between visual complexity and aesthetics: application to beauty prediction of photos. In *Computer vision-ECCV 2014 workshops: zurich, Switzerland, September 6-7 and 12, 2014, proceedings, part i 13* (pp. 20–34). Springer.

Van Geert, E., & Wagemans, J. (2020). Order, complexity, and aesthetic appreciation. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(2), 135.

Van Geert, E., & Wagemans, J. (2021). Order, complexity, and aesthetic preferences for neatly organized compositions. *Psychology of Aesthetics, Creativity, and the Arts*, *15*(3), 484.

Yu, H., & Winkler, S. (2013). Image complexity and spatial information. In *2013 fifth international workshop on quality of multimedia experience* (pp. 12–17). IEEE.