

Autonomously Communicating Conceptual Knowledge Through Visual Art

Derrall Heath, David Norton, Dan Ventura

Computer Science Department
Brigham Young University
Provo, UT 84602 USA
dheath@byu.edu, dnorton@byu.edu, ventura@cs.byu.edu

Abstract

In visual art, the communication of meaning or intent is an important part of eliciting an aesthetic experience in the viewer. Building on previous work, we present three additions to DARCI that enhances its ability to communicate concepts through the images it creates. The first addition is a model of semantic memory based on word associations for providing meaning to concepts. The second addition composes universal icons into a single image and renders the image to match an associated adjective. The third addition is a similarity metric that maintains recognizability while allowing for the introduction of artistic elements. We use an online survey to show that the system is successful at creating images that communicate concepts to human viewers.

Introduction

DARCI (Digital ARTist Communicating Intention) is a system for generating original images that convey meaning. The system is part of ongoing research in the subfield of computational creativity, and is inspired by other artistic image generating systems such as AARON (McCorduck 1991) and The Painting Fool (Colton 2011). Central to the design philosophy of DARCI is the notion that the communication of meaning in art is a necessary part of eliciting an aesthetic experience in the viewer (Csikzentmihályi and Robinson 1990). DARCI is unique from other computationally creative systems in that DARCI creates images that explicitly express a given concept.

DARCI is composed of two major subsystems, an *image analysis* component, and an *image generation* component. The image analysis component learns how to annotate images with adjectives by training a series of neural networks with labeled images. The specific inputs to these neural networks, called *appreciation networks*, are global features extracted from each image, including information about the general occurrence of color, lighting, and texture in the images (Norton, Heath, and Ventura 2010). The image generation component uses a genetic algorithm, governed partly by the analysis component, to render a *source image* to visually convey an adjective (Norton, Heath, and Ventura 2011). While often effective, excessive filtering and extreme parameters can leave the source image unrecognizable.

In this paper we introduce new capabilities to DARCI—primarily, the ability to produce original source images

rather than relying upon pre-existing, human-provided images. DARCI composes these original source images as a collage of iconic concepts in order to express a range of concepts beyond adjectives, similar to a recently introduced system for The Painting Fool that creates collages from the text of web documents (Krzeczkowska et al. 2010). However, in contrast to that system, ours creates collages from conceptual icons discovered with a semantic memory model. The resulting source images are then rendered according to an adjective discovered with this same semantic memory model. In order to preserve the content of the collages after rendering them, we introduce a variation on DARCI’s traditional image rendering technique. Figure 1 outlines the two major components and their interaction, including the new elements presented in this paper. By polling online volunteers, we show that with these additions, DARCI is capable of creating images that convey selected concepts while maintaining the aesthetics achieved with filters.

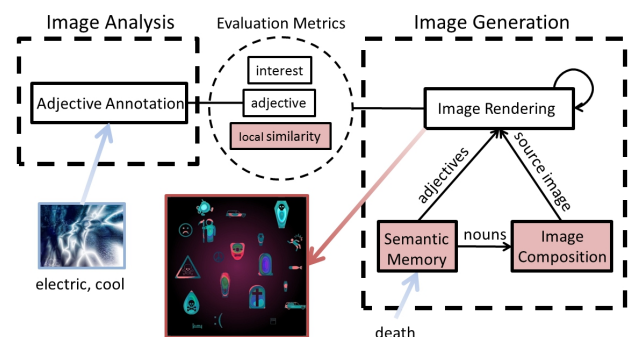


Figure 1: A diagram outlining the two major components of DARCI. *Image analysis* learns how to annotate new images with adjectives using a series of *appreciation networks* trained with labeled images. *Image generation* uses a *semantic memory* model to identify nouns and adjectives associated with a given concept. The nouns are composed into a source image that is rendered to reflect the adjectives, using a genetic algorithm that is governed by a set of evaluation metrics. The final product is an image that reflects the given concept. Additions from this paper are highlighted.

Methodology

Here we introduce the improvements to DARCI that enhance the system's capability to communicate intended meaning in an aesthetic fashion: a semantic memory model for broadening the range of concepts the system can communicate, an image composer for composing concrete representations of concepts into source images to be rendered, and a new metric for governing the evolution of the rendering process. We also describe an online survey that we use to evaluate the success of these additions.

Semantic Memory Model

In cognitive psychology, the term *semantic memory* refers to the memory of meaning and other concept-based knowledge that allows people to consciously recall general information about the world. It is often argued that creativity requires intention (and we are certainly in this camp). In this context, we mean creativity in communicating a concept, and at least one part of this can be accommodated by an internal knowledge of the concept (i.e., a semantic memory).

The question of what gives words (or concepts) meaning has been debated for years; however, it is commonly agreed that a word, at least in part, is given meaning by how the word is used in conjunction with other words (i.e., its context) (Erk 2010). Many computational models of semantic memory consist of building associations between words (Sun 2008; De Deyne and Storms 2008), and these word associations essentially form a large graph that is typically referred to as a *semantic network*. Associated words provide a level of meaning to a concept (word) and can be used to help convey its meaning.

Word associations are commonly acquired in one of two ways: from people and automatically by inferring them from a corpus. Here we describe a computational model of semantic memory that combines human free association norms with a simple corpus-based approach. The idea is to use the human word associations to capture general knowledge and then to fill in the gaps using the corpus method.

Lemmatization and Stop Words In gathering word associations, we use the standard practice of removing stop words and lemmatizing. The latter process is accomplished using WordNet's (Fellbaum 1998) database of word forms; it should be noted, however, that lemmatization with WordNet has its limits. For example, we cannot lemmatize a word across different parts of speech. As a result, words like 'redeem' and 'redeeming' will remain separate concepts because 'redeeming' could be the gerund form of the verb 'redeem' or it could be an adjective (i.e., the act of 'a redeeming quality').

Free Association Norms One of the most common means of gathering word associations from people is through *Free Association Norms* (FANs), which is done by asking hundreds of human volunteers to provide the first word that comes to mind when given a cue word. This technique is able to capture many different types of word associations including word co-ordination (pepper, salt), collocation (trash, can), super-ordination (insect, butterfly), synonymy (starving, hungry), and antonymy (good, bad). The association

strength between two words is simply a count of the number of volunteers that said the second word given the first word. FANs are considered to be one of the best methods for understanding how people, in general, associate words in their own minds (Nelson, McEvoy, and Schreiber 1998). In our model we use two preexisting databases of FANs: The Edinburgh Associative Thesaurus (Kiss et al. 1973) and the University of Florida's Word Association Norms (Nelson, McEvoy, and Schreiber 1998).

Note that in this model we consider word associations to be undirected. In other words, if word A is associated with word B , then word B is associated with word A . Hence, when we encounter data in which word A is a cue for word B and word B is also a cue for word A , we combine them into a single association pair by adding their respective association strengths. Between these two databases, there are a total of 19,327 unique words and 288,069 unique associations. We refer to these associations as *human data*.

Corpus Inferred Associations Discovering word associations from a corpus is typically accomplished using a family of techniques called *Vector Space Models* (Turney and Pantel 2010), which uses a matrix for keeping track of word counts either co-occurring with other words (term \times term matrix) or within each document (term \times document matrix).

One of the most popular vector space models is *Latent Semantic Analysis* (LSA) (Deerwester et al. 1990), based on the idea that similar words will appear in similar documents (or contexts). LSA builds a term \times document matrix from a corpus and then performs Singular Value Decomposition (SVD), which essentially reduces the large sparse matrix to a low-rank approximation of that matrix along with a set of vectors, each representing a word (as well as a set of vectors for each document). These vectors also represent points in semantic space, and the closer words are to each other in this space, the closer they are in meaning (and the stronger the association between words).

Another popular method is the *Hyperspace Analog to Language* (HAL) model (Lund and Burgess 1996). This model is based on the same idea as LSA, except the notion of context is reduced more locally to a word co-occurrence window of ± 10 words instead of an entire document. Thus, the HAL model builds a term \times term matrix of word co-occurrence counts from a corpus. HAL then uses the co-occurrence counts directly as vectors representing each word in semantic space. The size of the term \times term matrix is invariant to the size of the corpus and has been argued to be more congruent to human cognition than the term \times document matrix used in LSA (Wandmacher, Ovchinnikova, and Alexandrov 2008; Burgess 1998).

The corpus component of our model is constructed similarly to HAL but with some important differences. We restrict the model to the same number of unique words as the human-generated free associations, building a $19,327 \times 19,327$ (term \times term) co-occurrence matrix M using a co-occurrence window of ± 50 words. To account for the fact that common words will have generally higher co-occurrence counts, we scale these counts by weighting each element of the matrix by the inverse of the total frequency

of both words at each element. This is done by considering each element $M_{i,j}$, then adding the total number of occurrences of each word (i and j), subtracting out the value at $M_{i,j}$ (to avoid counting it twice), then dividing $M_{i,j}$ by this computed number, as follows:

$$M_{i,j} \leftarrow \frac{M_{i,j}}{(\sum_i M_{i,j} + \sum_j M_{i,j} - M_{i,j})} \quad (1)$$

The result could be a very small number, and therefore we then also normalize the values between 0 and 1.

For our corpus we use Wikipedia, as it is large, easily accessible, and covers a wide range of human knowledge (Denoyer and Gallinari 2006). Once the co-occurrence matrix is built from the entire text of Wikipedia, we use the weighted/normalized co-occurrence values themselves as association strengths between words. This approach works, since we only care about the strongest associations between words, and it allows us to reduce the number of irrelevant associations by ignoring any word pairs with a co-occurrence count less than some threshold. We chose a threshold of 100 (before weighting), which provides a good balance of producing a sufficient number of associations, while reducing the number of irrelevant associations. When looking up a particular word, we return the top n other words with the highest weighted/normalized co-occurrence values. This method, which we will call *corpus data* from now on, gives a total of 4,908,352 unique associations.

Combining Word Associations Since each source (human and corpus) provide different types of word associations, a combination of these methods into a single model has the potential to take advantage of the strengths of each method. The hypothesis is that the combined model will better communicate meaning to a person than either model individually because it presents a wider range of associations.

Our method merges the two separate databases into a single database before querying it for associations. This method assumes that the human data contains more valuable word associations than the corpus data because the human data is typically used as the gold standard in the literature. However, the corpus data does contain some valuable associations not present in the human data. The idea is to add the top n associations for each word from the corpus data to the human data but to weight the association strength low. This is beneficial for two reasons. First, if there are any associations that overlap, adding them again will strengthen the association in the combined database. Second, new associations not present in the human data will be added to the combined database and provide a greater variety of word associations. We keep the association strength low because we want the corpus data to reinforce, but not dominate, the human data.

To do this, we first copy all word associations from the human data to the combined database. Next, let W be the set of all 19,327 unique words, let $A_{i,n} \subseteq W$ be the set of the top n words associated with word $i \in W$ from the corpus data, let $score_{i,j}$ be the association strength between words i and j from the corpus data, let max_i be the maximum

association score present in the human data for word i , and let θ be a weight parameter. Now for each $i \in W$ and for each $j \in A_{i,n}$, the new association score between words i and j is computed as follows:

$$score_{i,j} \leftarrow (max_i \cdot \theta) \cdot score_{i,j} \quad (2)$$

This equation scales $score_{i,j}$ (which is already normalized) to lie between 0 and a certain percentage (θ) of max_i . The n associated words from the corpus are then added to the combined database with the updated scores. If the word pair is already in the database, then the updated score is added to the score already present. For the results presented in this paper we use $n = 20$ and $\theta = 0.2$, which were determined based on preliminary experiments. After the merge, the combined database contains 443,609 associations.

Image Composer

The semantic memory model can be considered to represent the meaning of a word as a (weighted) collection of other words. DARCI effectively makes use of this collection as a decomposition of a (high-level) concept into simpler concepts that together represent the whole, the idea being that in many cases, if a (sub)concept is simple enough, it can be represented visually with a single icon (e.g., the concept ‘rock’ can be visually represented with a picture of a ‘rock’). Given such collection of iconic concepts, DARCI composes their visual representations (icons) into a single image. The image is then rendered to match some adjective associated with the original (collective) concept.

To represent these “simple enough” concepts, DARCI makes use of a collection of icons provided by *The Noun Project*, whose goal is to build a repository of symbols/icons that can be used as a visual language (Thomas et al. 2013). The icons are intended to be simple visual representations of any noun and are published by various artists under the Creative Commons license. Currently, The Noun Project provides 6,334 icons (each 420×420 pixels) representing 2,535 unique nouns and is constantly growing.

When given a concept, DARCI first uses the semantic memory model to retrieve all words associated with the given concept, including itself. These word associations are filtered by returning only nouns for which DARCI has icons and adjectives for which DARCI has appreciation networks. The nouns are sorted by association strength and the top 15 are kept. For each noun, multiple icons are usually available and one or two of these icons are chosen at random to create a set of icons for use in composing the image.

The icons in the set are scaled to between 25% and 100% of their original size according to their association strength rank. Let I be the set of icons, and let $r : I \rightarrow [0, |I| - 1]$ be the rank of icon $i \in I$, where the icon with rank 0 corresponds to the noun with the highest association strength. Finally, let ϕ_i be the scaling factor for icon i , which is computed as follows:

$$\phi_i \leftarrow 1 - \frac{0.75}{|I| - 1} r(i) \quad (3)$$

An initial blank white image of size 2000×2000 pixels is created and the set of scaled icons are drawn onto the blank

image at random locations, the only constraints being that no icons are allowed to overlap and no icons are allowed to extend beyond the border of the image. The result is a collage of icons that represents the original concept. DARCI then randomly selects an adjective from the set returned by the semantic memory model weighted by each adjective’s association strength. DARCI uses its adjective rendering component, described in prior work, to render the collage image according to the selected adjective (Norton, Heath, and Ventura 2011; 2013; Heath, Norton, and Ventura 2013). The final image will both be artistic and in some way communicate the concept to the viewer. Figure 1 shows how this process is incorporated into the full system.

Similarity Metric

To render an image, DARCI uses a genetic algorithm to discover a combination of filters that will render a source image (in this case, the collage) to match a specified adjective. The fitness function for this process combines an *adjective metric* and an *interest metric*. The former measures how effectively a potential rendering, or *phenotype*, communicates the adjective, and the latter measures the “difference” between the phenotype and the source image. Both metrics use only global image features and so fail to capture important local image properties correlated with image content.

In this paper we introduce a third metric, *similarity*, that borrows from the growing research on *bag-of-visual-word* models (Csurka et al. 2004; Sivic et al. 2005) to analyze local features, rather than global ones. Typically, these interest points are those points in an image that are the most surprising, or said another way, the least predictable. After an interest point is identified, it is described with a vector of features obtained by analyzing the region surrounding the point. *Visual words* are quantized local image features. A dictionary of visual words is defined for a domain by extracting local interest points from a large number of representative images and then clustering them (typically with *k*-means) by their features into *n* clusters, where *n* is the desired dictionary size. With this dictionary, visual words can be extracted from any image by determining which clusters the image’s local interest points belong. A bag-of-visual-words for the image can then be created by organizing the visual word counts for the image into a fixed vector. This model is analogous to the bag-of-words construct for text documents in natural language processing.

For the new *similarity metric*, we first create a bag-of-visual-words for the source image and each phenotype, and then calculate the Euclidean distance between these two vectors. This metric has the effect of measuring the number of interest points that coincide between the two images.

We use the standard SURF (Speeded-Up Robust Features) detector and descriptor to extract interest points and their features from images (Bay et al. 2008). SURF quickly identifies interest points using an approximation of the difference of Gaussians function, which will often identify corners and distinct edges within images. To describe each interest point, SURF first assigns an orientation to the interest point based on surrounding gradients. Then, relative to this orientation, SURF creates a 64 element feature vector by summing both

the values and magnitudes of Haar wavelet responses in the horizontal and vertical directions for each square of a four by four grid centered on the point.

We build our visual word dictionary by extracting these SURF features from the database of universal icons mentioned previously. The 6334 icons result in more than two hundred thousand interest points which are then clustered into a dictionary of 1000 visual words using Elkan *k*-means (Elkan 2003). Once the Euclidean distance, *d*, between the source image’s and the phenotype’s bags-of-visual-words is calculated, the metric, *S*, is calculated to provide a value between 0 and 1 as follows:

$$S = \text{MAX}\left(\frac{d}{100}, 1\right)$$

where the constant 100 was chosen empirically.

Online Survey

Since our ultimate goal is a system that can create images that both communicate intention and are aesthetically interesting, we have developed a survey to test our most recent attempts at conveying concepts while rendering images that are perceived as creative.

The survey asks users to evaluate images generated for ten concepts across three rendering techniques. The ten concepts were chosen to cover a variety of abstract and concrete topics. The abstract concepts are ‘adventure’, ‘love’, ‘music’, ‘religion’, and ‘war’. The concrete concepts are ‘bear’, ‘cheese’, ‘computer’, ‘fire’, and ‘garden’.

We refer to the three rendering techniques as *unrendered*, *traditional*, and *advanced*. For *unrendered*, no rendering is applied—these are the plain collages. For the other two techniques, the images are rendered using one of two fitness functions to govern the genetic algorithm. For *traditional*, the fitness function is the average of the adjective and interest metrics. For *advanced* rendering, the new similarity metric is added. Here the adjective metric is weighted by 0.5, while the interest and similarity metrics are each weighted by 0.25. For each rendering technique and image, DARCI returned the 40 highest ranking images discovered over a period of 90 generations. We then selected from the pools of 40 for each concept and technique, the image that we felt best conveyed the intended concept while appearing aesthetically interesting. An example image that we selected from each rendering technique can be seen in Figure 2.

To query the users about each image, we followed the survey template that we developed previously to study the perceived creativity of images rendered with different adjectives (Norton, Heath, and Ventura 2013). In this study, we presented users with six five-point Likert items (Likert 1932) per image; volunteers were asked how strongly they agreed or disagreed (on a five point scale) with each statement as it pertained to one of DARCI’s images. The six statements we used were (abbreviation of item in parentheses):

- I like the image. (*like*)
- I think the image is novel. (*novel*)
- I would use the image as a desktop wallpaper. (*wallpaper*)
- Prior to this survey, I have never seen an image like this one. (*never seen*)
- I think the image would be difficult to create. (*difficult*)
- I think the image is creative. (*creative*)

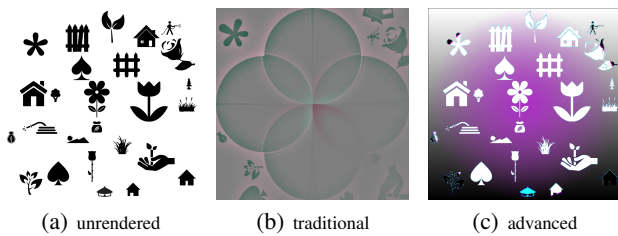


Figure 2: Example images¹ for the three rendering techniques representing the concept ‘garden’.

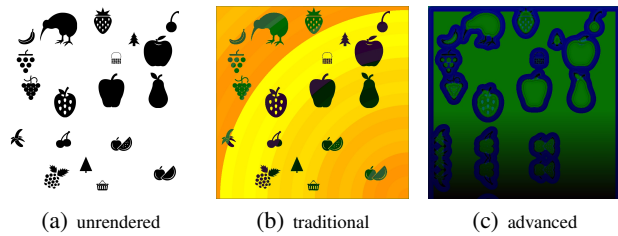


Figure 3: Example dummy images² for the concept ‘water’ that appeared in the survey for the indicated rendering techniques.

In previous work, we showed that the first five statements correlated strongly with the sixth, “I think the image is creative” (Norton, Heath, and Ventura 2013), justifying this test as an accurate evaluation of an image’s subjective creativity. In this paper, we use the same six Likert items and add a seventh to determine how effective the images are at conveying their intended concept:

I think the image represents the concept of “_____.” (*concept*)

To avoid fatigue, volunteers were only presented with images from one of the three rendering techniques mentioned previously. The technique was chosen randomly and then the images were presented to the user in a random order. To help gauge the results, three dummy images were introduced into the survey for each technique. These dummy images were created for arbitrary concepts and then assigned different arbitrary concepts for the survey so that the image contents would not match their label. Unfiltered dummy collages were added to the unrendered set of images, while traditionally rendered versions were added to the traditional and advanced sets of images. The three concepts used to generate the dummy images were: ‘alien’, ‘fruit’, and ‘ice’. The three concepts that were used to describe these images in the survey were respectively: ‘restaurant’, ‘water’, and ‘freedom’. To avoid confusion, from here on we will always refer to these dummy images by their description word. The

¹The original icons used for the images in Figure 2 were designed by Adam Zubin, Birdie Brain, Evan Caughey, Rachel Fisher, Prerak Patel, Randall Barriga, dsathiyaraj, Jeremy Bristol, Andrew Fortnum, Markus Koltringer, Bryn MacKenzie, Hernan Schlosman, Maurizio Pedrazzoli, Mike Endale, George Agpoon, and Jacob Eckert of The Noun Project.

²The original icons used for the images in Figure 3 were designed by Alessandro Suraci, Anna Weiss, Riziki P.M.G. Nielsen, Stefano Bertoni, Paulo Volkova, James Pellizzi, Christian Michael Witternigg, Dan Christopher, Jayme Davis, Mathies Janssen, Pavel Nikandrov, and Luis Prado of The Noun Project.

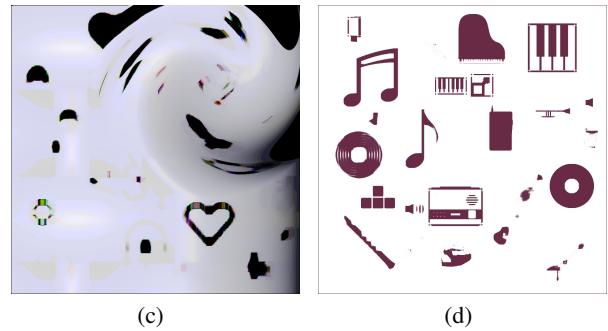
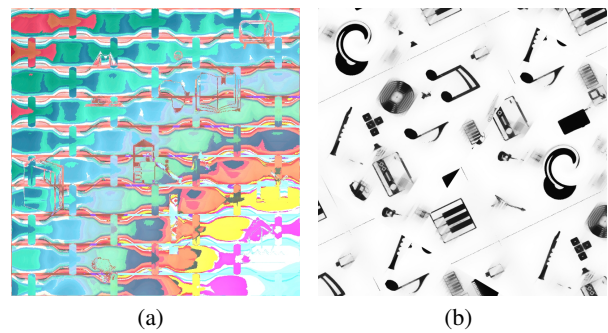


Figure 4: The images³ that were rated the highest on average for each statement. Image (a) is the advanced rendering of ‘adventure’ and was rated highest for *like*, *novel*, *difficult*, and *creative*. Image (b) is the traditional rendering of ‘music’ and was rated highest for *wallpaper*. Image (c) is the advanced rendering of ‘love’ and was rated highest for *never seen*. Image (d) is the advanced rendering of ‘music’ and was rated highest for *concept*.

dummy images for the concept of ‘water’ are shown in Figure 3. In total, each volunteer was presented with 13 images.

Results

A total of 119 anonymous individuals participated in the on-line survey. Volunteers could quit the survey at anytime, thus not evaluating all 13 images. Each person evaluated an average of 9 images and each image was evaluated by an average of 27 people. The highest and lowest rated images for each question can be seen in Figures 4 and 5 respectively.

The three dummy images for each rendering technique are used as a baseline for the concept statement. The results of the dummy images versus the valid images are shown in Figure 6. The average concept rating for the valid images is significantly better than the dummy images, which shows that the intended meaning is successfully conveyed to human viewers more reliably than an arbitrary image. These results confirm that the intelligent use of iconic concepts is beneficial for the visual communication of meaning. Further, it is suggestive that the ratings for the other statements are generally lower for the dummy images than for the valid

³The original icons used for the images in Figure 4 were designed by Oxana Devochkina, Kenneth Von Alt, Paul te Kortschot, Marvin Kutscha, James Fenton, Camilo Villegas, Gustavo Perez Rangel, and Anuar Zhumaev of The Noun Project.

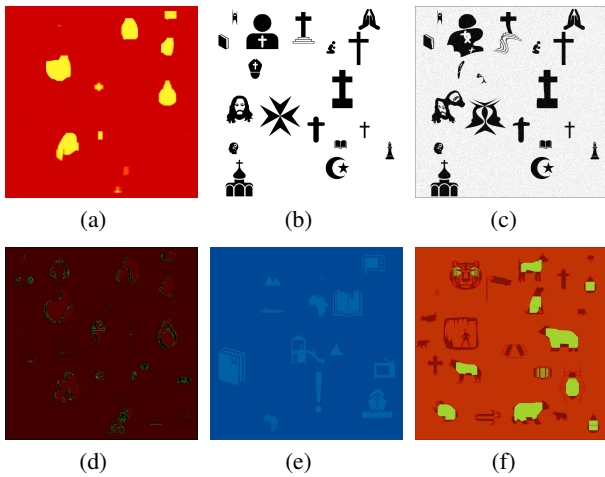


Figure 5: The images⁴ that were rated the lowest on average for each statement. Image (a) is the advanced rendering of ‘fire’ and was rated lowest for *difficult* and *creative*. Images (b) and (c) are the unrendered and advanced version of ‘religion’ and were rated lowest for *neverseen* and *wallpaper* respectively. Images (d), (e), and (f) are the traditional renderings of ‘fire’, ‘adventure’, and ‘bear’, respectively, and were rated lowest for *like*, *novel*, and *concept* respectively.

images. Since the the dummy images were created for a different concept than the one which they purport to convey in the survey, this may be taken as evidence that successful conceptual or intentional communication is an important factor for the attribution of creativity.

The results of the three rendering techniques (unrendered, traditional, and advanced) for all seven statements are shown in Figure 7. The unrendered images are generally the most successful at communicating the intended concepts. This is likely because the objects/icons in the unrendered images are left undisturbed and are therefore more clear and discernible, requiring the least perceptual effort by the viewer. The rendered images (traditional and advanced) often distort the icons in ways that make them less cohesive and less discernible and can thus obfuscate the intended meaning. The trade-off, of course, is that the unrendered images are generally considered less likable, less novel, and less creative than the rendered images. The advanced images are generally considered more novel and creative than the traditional images, but the traditional images are liked slightly more. The advanced images also convey the intended meaning more reliably than the traditional images, which indicates that the similarity metric is finding a better balance between adding artistic elements and maintaining icon recognizability.

The difference between the traditional and advanced rendering was minimized by the fact that we selected the image

⁴The original icons used for the images in Figure 5 were designed by Melissa Little, Dan Codyre, Carson Wittenberg, Kenneth Von Alt, Nicole Kathryn Griffing, Jenifer Cabrera, Renee Ramsey-Passmore, Ben Rex Furneaux, Factorio.us collective, Anuar Zhumaev, Luis Prado, Ahmed Hamzawy, Michael Rowe, Matthias Schmidt, Jule Steffen, Monika Ciapala, Bru Rakoto, Patrick Trouv, Adam Heller, Marco Acri, Mehmet Yavuz, Allison Dominguez, Dan Christopher, Nicholas Burroughs, Rodney Lobos, and Norman Ying of The Noun Project.

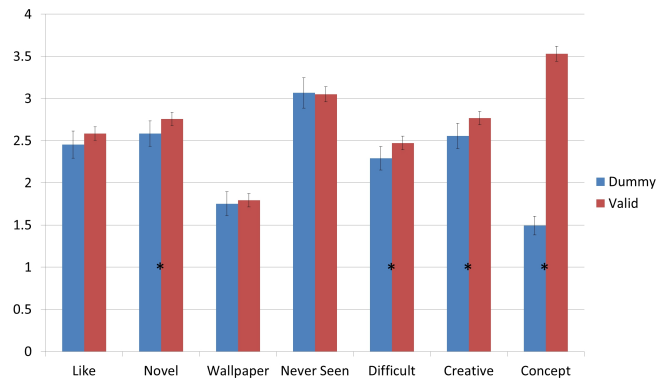


Figure 6: The average rating from the online survey for all seven statements comparing the dummy images with the valid images. The valid images were more successful at conveying the intended concept than the dummy images by a significant margin. Results marked with an asterisk (*) indicate statistical significance using the two tailed independent *t*-test. The lines at the top of each bar show the 95% confidence interval for each value. The sample sizes for dummy and valid images are 251 and 818 respectively.

(out of DARCI’s top 40) from each group that best conveyed the concept while also being aesthetically interesting. Out of all the traditional images, 39% had at least one recognizable icon, while 74% of the advanced images had at least one recognizable icon. This difference demonstrates that the new similarity metric helps to preserve the icons and provides a greater selection of good images from which to choose, which is consistent with the results of the survey. For comparison, Figure 8 shows some example images (both traditional and advanced) that were not chosen for the survey.

The results comparing the abstract concepts with the concrete concepts are shown in Figure 9. For all seven statements, the abstract concepts are, on average, rated higher than the concrete concepts. One possible reason for this is that concrete concepts are not easily decomposed into a collection of iconic concepts because, being concrete, they are more likely to be iconic themselves. For concrete concepts, the nouns returned by the semantic memory model are usually other related concrete concepts, and it becomes difficult to tell which object is the concept in question. For example, the concept ‘bear’ returns nouns like ‘cave’, ‘tiger’, ‘forest’, and ‘wolf’, which are all related, but don’t provide much indication that the intended concept is ‘bear’. A person might be inclined to generalize to a concept such as ‘wildlife’. Another possible reason why abstract concepts result in better survey results than do concrete concepts is because abstract concepts allow a wider range of interpretation and are generally more interesting. For example, the concept ‘cheese’ would generally be considered straightforward to most people, while the concept ‘love’ could have variable meanings to different people in different circumstances. Hence, the

⁵The original icons used for the images in Figure 8 are the same as those used in Figures 4 and 5 with attribution to the same designers.

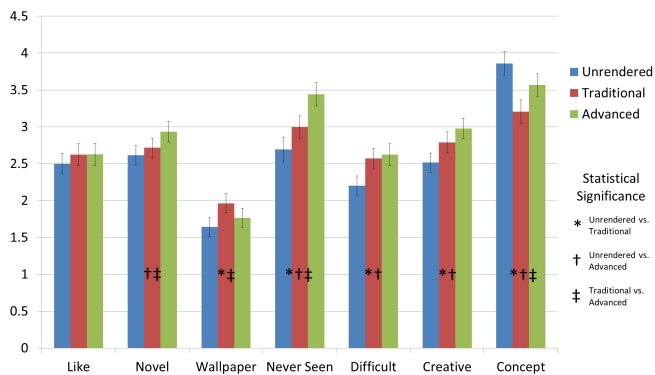


Figure 7: The average rating from the online survey for all seven statements comparing the three rendering techniques. The unrendered technique is most successful at representing the concept, while the advanced technique is generally considered more novel and creative. Statistical significance was calculated using the two tailed independent t -test. The lines at the top of each bar show the 95% confidence interval for each value. The sample sizes for the unrendered, traditional, and advanced techniques are 256, 285, and 277 respectively.

images generated for abstract concepts are generally considered more likable, more novel, and more creative than the concrete images.

Conclusions and Future Work

We have presented three additions to the computer system, DARCI, that enhance the system’s ability to communicate specified concepts through the images it creates. The first addition is a model of semantic memory that provides conceptual knowledge necessary for determining how to compose and render an image by allowing the system to make decisions and reason (in a limited manner) about common world knowledge. The second addition uses the word associations from a semantic memory model to retrieve conceptual icons and composes them into a single image, which is then rendered in the manner of an associated adjective. The third addition is a new similarity metric used during the adjective rendering phase that preserves the discernibility of the icons while allowing for the introduction of artistic elements.

We used an online survey to evaluate the system and show that DARCI is significantly better at expressing the meaning of concepts through the images it creates than an arbitrary image. We show that the new similarity metric allows DARCI to find a better balance between adding interesting artistic qualities and keeping the icons/objects recognizable. We show that using word associations and universal icons in an intelligent way is beneficial for conveying meaning to human viewers. Finally, we show that there is some degree of correlation between how well an image communicates the intended concept and how well liked, how novel, and how creative the image is considered to be. To further illustrate DARCI’s potential, Figure 10 shows additional images encountered during various experiments with DARCI that we

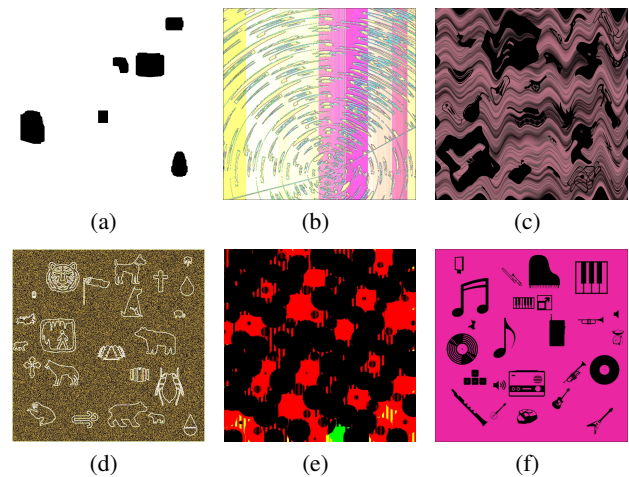


Figure 8: Sample images⁵ that were not chosen for the online survey. Images (a), (b), and (c) are traditional renderings of ‘adventure’, ‘love’, and ‘war’ respectively. Images (d), (e), and (f) are advanced renderings of ‘bear’, ‘fire’, and ‘music’ respectively.

thought were particularly interesting.

In future research we plan to do a direct comparison of the images created by DARCI with images created by human artists and to further investigate how semantic memory contributes to the creative process. We plan to improve the semantic memory model by going beyond word-to-word associations and building associations between words and other objects (such as images). This will require expanding DARCI’s image analysis capability to include some level of image noun annotation. The similarity metric presented in this paper is a step in that direction. An improved semantic memory model could also help enable DARCI to discover its own topics (i.e., find its own inspiration) and to compose icons together in more meaningful ways, by intentional choice of absolute and relative icon placement, for example.

References

- Bay, H.; Ess, A.; Tuytelaars, T.; and Gool, L. V. 2008. Speeded-up robust features (SURF). *Computer Vision and Image Understanding* 110:346–359.
- Burgess, C. 1998. From simple associations to the building blocks of language: Modeling meaning in memory with the HAL model. *Behavior Research Methods, Instruments, & Computers* 30:188–198.
- Colton, S. 2011. The Painting Fool: Stories from building an automated painter. In McCormack, J., and d’Inverno, M., eds., *Computers and Creativity*. Springer-Verlag.
- Csikzentmihályi, M., and Robinson, R. E. 1990. *The Art of Seeing*. The J. Paul Getty Trust Office of Publications.
- Csurka, G.; Dance, C. R.; Fan, L.; Willamowski, J.; and Bray, C. 2004. Visual categorization with bags of keypoints.

⁵The original icons used for the images in Figure 10 were designed by Alfredo Astort, Simon Child, Samuel Eidam, and Jonathan Keating of The Noun Project.

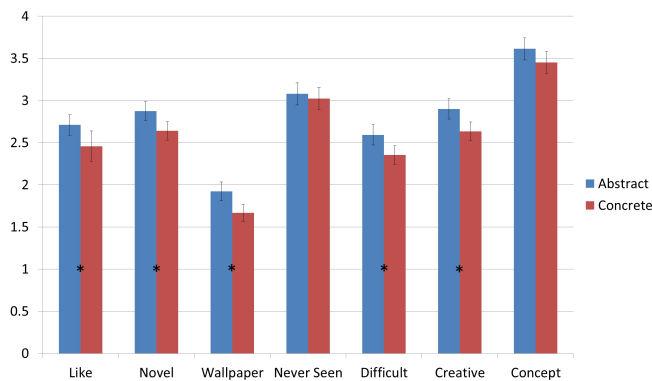


Figure 9: The average rating from the online survey for all seven statements comparing the abstract concepts with the concrete concepts. The abstract concepts generally received higher ratings for all seven statements. Results marked with an asterix (*) indicate statistical significance using the two tailed independent *t*-test. The lines at the top of each bar show the 95% confidence interval for each value. The sample sizes for abstract and concrete concepts are 410 and 408 respectively.

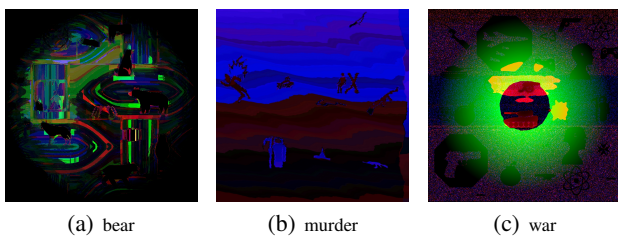


Figure 10: Notable images⁶ rendered by DARCI during various experiments and trials.

In *Proceedings of the Workshop on Statistical Learning in Computer Vision*, 1–22.

De Deyne, S., and Storms, G. 2008. Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods* 40(1):198–205.

Deerwester, S.; Dumais, S. T.; Furnas, G. W.; Landauer, T. K.; and Harshman, R. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41(6):391–407.

Denoyer, L., and Gallinari, P. 2006. The Wikipedia XML corpus. In *INEX Workshop Pre-Proceedings*, 367–372.

Elkan, C. 2003. Using the triangle inequality to accelerate *k*-means. In *Proceedings of the Twentieth International Conference on Machine Learning*, 147–153.

Erk, K. 2010. What is word meaning, really?: (and how can distributional models help us describe it?). In *Proceedings of the 2010 Workshop on GEometrical Models of Natural Language Semantics*, 17–26. Stroudsburg, PA, USA: Association for Computational Linguistics.

Fellbaum, C., ed. 1998. *WordNet: An Electronic Lexical Database*. The MIT Press.

Heath, D.; Norton, D.; and Ventura, D. 2013. Conveying semantics through visual metaphor. *ACM Transactions of Intelligent Systems and Technology, to appear*.

Kiss, G. R.; Armstrong, C.; Milroy, R.; and Piper, J. 1973. An associative thesaurus of English and its computer analysis. In Aitkin, A. J.; Bailey, R. W.; and Hamilton-Smith, N., eds., *The Computer and Literary Studies*. Edinburgh, UK: University Press.

Krzeczowska, A.; El-Hage, J.; Colton, S.; and Clark, S. 2010. Automated collage generation — with intent. In *Proceedings of the 1st International Conference on Computational Creativity*, 36–40.

Likert, R. 1932. A technique for the measurement of attitudes. *Archives of Psychology* 22(140):1–55.

Lund, K., and Burgess, C. 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers* 28:203–208.

McCorduck, P. 1991. *AARON's Code: Meta-Art, Artificial Intelligence, and the Work of Harold Cohen*. W. H. Freeman & Co.

Nelson, D. L.; McEvoy, C. L.; and Schreiber, T. A. 1998. The University of South Florida word association, rhyme, and word fragment norms. <http://www.usf.edu/FreeAssociation/>.

Norton, D.; Heath, D.; and Ventura, D. 2010. Establishing appreciation in a creative system. In *Proceedings of the 1st International Conference on Computational Creativity*, 26–35.

Norton, D.; Heath, D.; and Ventura, D. 2011. Autonomously creating quality images. In *Proceedings of the 2nd International Conference on Computational Creativity*, 10–15.

Norton, D.; Heath, D.; and Ventura, D. 2013. Finding creativity in an artificial artist. *Journal of Creative Behavior, to appear*.

Sivic, J.; Russell, B. C.; Efros, A. A.; Zisserman, A.; and Freeman, W. T. 2005. Discovering objects and their location in images. *International Journal of Computer Vision* 1:370–377.

Sun, R. 2008. *The Cambridge Handbook of Computational Psychology*. New York, NY, USA: Cambridge University Press, 1st edition.

Thomas, S.; Boatman, E.; Polyakov, S.; Mumenthaler, J.; and Wolff, C. 2013. The noun project. <http://thenounproject.com>.

Turney, P. D., and Pantel, P. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37:141–188.

Wandmacher, T.; Ovchinnikova, E.; and Alexandrov, T. 2008. Does latent semantic analysis reflect human associations? In *Proceedings of the ESSLLI Workshop on Distributional Lexical Semantics*, 63–70.