Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Creating Images by Learning Image Semantics Using Vector Space Models

Derrall Heath and **Dan Ventura**
Computer Science Department
Brigham Young University
Provo, UT 84602 USA
dheath@byu.edu, ventura@cs.byu.edu

Abstract

When dealing with images and semantics, most computational systems attempt to automatically extract meaning from images. Here we attempt to go the other direction and autonomously create images that communicate concepts. We present an enhanced semantic model that is used to generate novel images that convey meaning. We employ a vector space model and a large corpus to learn vector representations of words and then train the semantic model to predict word vectors that could describe a given image. Once trained, the model autonomously guides the process of rendering images that convey particular concepts. A significant contribution is that, because of the semantic associations encoded in these word vectors, we can also render images that convey concepts on which the model was not explicitly trained. We evaluate the semantic model with an image clustering technique and demonstrate that the model is successful in creating images that communicate semantic relationships.

Introduction

When considering the relationship between images and meaning (or semantics), most computational systems focus on extracting meaning from images. For example, image annotation (Wang 2011) and content-based image retrieval (Liu et al. 2007) are two major topics in computer vision whose goal is to automatically understand the semantics within images. Here we focus on going the other direction, that is to *generate* images based on semantics.

There are few systems we know of that attempt to autonomously generate images that communicate meaning. The WordsEye system tries to generate 3D scenes based on written descriptions (Coyne and Sproat 2001). The Story Picture Engine (Joshi, Wang, and Li 2006) and the Text-to-Picture Synthesis System (Zhu et al. 2007) are both systems built to do automatic text illustration (i.e., to visually tell a story or to graphically communicate the gist of text). AARON (McCorduck 1991) and The Painting Fool (Colton 2011) are both systems designed to autonomously create visual art in ways meaningful to human viewers.

Our own system, DARCI, is designed to create novel, artistic images that explicitly express a given concept (Norton, Heath, and Ventura 2015). Central to the design philosophy of DARCI is the notion that the communication of meaning in visual art is a necessary part of eliciting an aesthetic experience in the viewer (Csíkzentmihályi and Robinson 1990). In this paper we present a sophisticated semantic model that allows DARCI to internally represent the meaning of concepts and to express these concepts through images in novel ways.

It is commonly agreed that a word (or concept), at least in part, is given meaning by how the word is used in conjunction with other words (i.e., its context) (Landauer and Dumais 1997; Erk 2010). Vector Space Models (VSMs) are common methods for automatically learning vector representations of word meaning from a large corpus (Turney and Pantel 2010). These models are based on the idea that similar words will occur in similar contexts and words that are often associated together will often co-occur close together. These models reduce words to a vector representation that can be compared to other word vectors.

VSMs have been successfully used on a variety of tasks such as information retrieval (Salton 1971), multiple choice vocabulary tests (Denhière and Lemaire 2004), multiple choice synonym questions from the TOEFL test (Rapp 2003), multiple choice analogy questions from the SAT test (Turney 2006), and object recognition systems (Frome et al. 2013). Additionally, an approach called the ACI (Associative Conceptual Imagination) framework has recently been proposed as a way to use VSMs for imaginative and generative tasks (Heath, Dennis, and Ventura 2015).

We apply the ACI framework to DARCI by incorporating a VSM and building a visual semantic model that uses a large neural network to learn associations between low-level image features and adjective vectors from the VSM. This visual semantic model allows DARCI to create images that convey the meaning of adjectives to the viewer. It also allows DARCI to take advantage of the semantic structure between words and render images according to adjectives on which it was never explicitly trained. For example, DARCI could be trained on 'scary' and 'dark' images, but not 'creepy' images. DARCI could then "imagine" what a 'creepy' image would look like because 'creepy' is similar in meaning to 'scary' and 'dark'. Even higher level concepts (e.g., 'love', 'freedom') can be partially expressed through the images DARCI renders.

Clustering techniques have been previously developed to

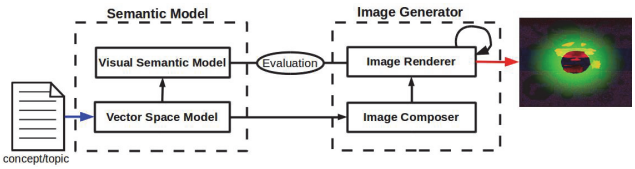


Figure 1: The two major components of DARCI. The *semantic model* first learns vector representations of words by analyzing a corpus (vector space model). The visual semantic model then learns to predict these word vectors using a neural network trained with labeled images. The *image generator* uses the vector space model to identify other words associated with a given concept. The nouns are composed into a source image (image composer) that is rendered to convey the original concept using a genetic algorithm (image renderer) that is governed in part by the visual semantic model. The final product is an image that reflects the given concept.

measure how well rendered images convey descriptive concepts (Heath, Norton, and Ventura 2014). We apply these clustering methods here and show that the new semantic model successfully enables DARCI to render images that convey a larger variety of concepts in ways that accurately reflect their semantic relationships.

Methodology

DARCI is composed of two major subsystems, a *semantic model* and an *image generator* as shown in Figure 1. We outline in detail the semantic model, which includes a state-of-the-art VSM that learns semantic relationships between words, and an artificial neural network that does multi-target regression to associate image features with the word vectors inferred from the VSM. We then describe the image generator and how it interacts with the semantic model to create meaningful images. Note that the image generator is not the focus of this paper, and further details, including extensive evaluation, can be found in prior work (Norton, Heath, and Ventura 2015).

Semantic Model

In order for DARCI to express semantic information through pictures, it must first have its own semantic knowledge that can influence the images it creates. Our goal is to leverage semantic information gained through written text and transfer it to the task of meaningful image generation.

Vector Space Model We use a state-of-the-art VSM, called the *skip-gram model* (Mikolov et al. 2013). The skip-gram model is a neural architecture that analyses a large corpus and learns to predict the surrounding words given a current word. During training, the skip-gram model consequently learns vector representations for each word, which encode semantic information. Words similar in meaning will have vectors that are close to each other in "vector space". These word vectors capture other interesting semantic relationships that are consistent with arithmetic operations. For example, $vector(\text{“}king\text{”}) - vector(\text{“}man\text{”}) + vector(\text{“}woman\text{”})$ results in a vector that is closest to $vector(\text{“}queen\text{”})$.

These semantic vectors allow DARCI to find concepts related to a given word and to assess the similarity in meaning between words, which will aid DARCI in creating meaningful images. We use a publicly available implementation of the skip-gram model[1] and a lemmatized Wikipedia corpus to learn the word vectors (Denoyer and Gallinari 2006). The skip-gram implementation is used with out-of-the-box parameters except for the vector size, which is set to 300. The choice of 300 provides a balance between encoding enough semantic information to be useful and ease of prediction when associating the vectors with image features.

Visual Semantic Model In order for DARCI to leverage the word vectors for image creation, it must learn to associate image qualities with the semantic vectors. Currently, we limit the associated words to vectors representing adjectives and use a neural network model to predict the adjective vector for a given image.

We maintain a dataset of approximately 15,000 images that have either been explicitly hand labeled or automatically retrieved through Google image search. Once an adjective has enough labeled images (20 positive and 20 negative), we begin learning that adjective. As of this paper, there are 145 adjectives that meet this threshold. We extract from each image 51 global and local features representing attributes like color, lighting, texture, and local interest points, and have been shown to work well for emotional and descriptive labels (Norton, Heath, and Ventura 2016).

We train two separate neural networks, one with the positively labeled images, and one with the negatively labeled images. The positive network tries to predict what adjective an image IS, while the negative network tries to predict what adjective an image IS NOT. These networks learn to predict the appropriate adjective *vector* given an image. We treat this as a multi-target regression problem and initialize each neural network with 300 output nodes (one for each vector element). The inputs are the 51 image features, the hidden layer is non-linear (sigmoid), and the output layer is linear. The parameters for the neural networks were determined through experimentation (see the Evaluation Section for the metrics used) and include a learning rate of 0.01, a momentum of 0.1, and 100 hidden nodes. We use standard back-propagation with drop-out regularization to initially train the weights. Since the output layer is linear, we improved each model by solving for the least-squares solution as the final training step.

Figure 2 shows how the networks are used to determine how well an image matches an adjective. Let $\vec{v}_p$ and $\vec{v}_n$ be the vectors predicted by the positive and negative networks, respectively. Let $\vec{v}_a$ be the vector for adjective a from the VSM and let $sim(\vec{v}_1, \vec{v}_2)$ compute the cosine similarity between two vectors. Given an image, we can compute its score for a particular adjective using the following formula:

$$score = \frac{(sim(\vec{v}_p, \vec{v}_a) - sim(\vec{v}_n, \vec{v}_a)) + 1.0}{2.0} \quad (1)$$

[1] https://code.google.com/p/word2vec/

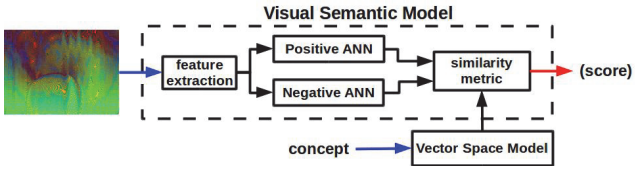


Figure 2: This diagram illustrates how the visual semantic model determines to what degree an image matches a given concept. It first extracts features from the image which are passed to both the positive neural network and the negative neural network. The word vector for the given concept is retrieved from the vector space model and compared via cosine similarity to the predicted vectors from the two neural networks. The similarity scores are combined and normalized for an overall score.

Learning to predict an adjective's vector is a harder task than learning to predict the adjective directly and, thus, introduces a few trade-offs. First, labeling images with adjectives is a multi-label classification problem (i.e., an image can be described by more than one valid adjective) and our new model can only predict one vector at a time, while normally each adjective could be predicted independently. The second trade-off is that our visual semantic model is predicting a 300 dimensional vector and has to account for every adjective. This means that it may not predict the 145 adjectives as accurately as using separate models for each adjective. The main advantage of learning the vectors, however, is that we can do zero-shot prediction. In other words, it is not limited to the 145 adjectives for which it was explicitly trained and can predict vectors for any adjective. The model can assess, how 'glad' an image is even if it has never seen a 'glad' picture because the semantic relationships of many adjectives are encoded in the vectors.

Image Generator

With the vector space and visual semantic models in place, DARCI can now produce images. Figure 1 shows how this process works. First, a concept/word/topic is given to the system and the VSM finds semantically related concepts. DARCI effectively makes use of these word associations 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 a collection of iconic concepts, DARCI composes their visual representations (icons) into a single image. This *source* image is then passed to the image renderer, which uses a genetic algorithm to render the image in an artistic way that conveys the meaning of the original concept. During rendering, the visual semantic model acts as the fitness function to guide the rendering process.

For example, suppose the original concept given to DARCI was 'war'. The vector space model would send related words like 'soldier', 'army', 'conflict', and 'battle' to the image composer. The resulting source image would be some composition of simple iconic images of the related words. The image renderer would then render this source image according to the visual semantic model. In this case the visual semantic model is telling the image renderer to create images that are close to the semantic vector for 'war'. However, since the visual semantic model was only trained on the 145 adjectives, this results in a rendering based on the adjectives that are semantically related to 'war' (in this case, 'bloody', 'violent', 'lonely', etc).

DARCI can also forgo the image composer and go straight to the image renderer, in which case the image produced will be an *abstract* rendering of the given concept. The user could also provide DARCI a source image of their own, like a photograph, and DARCI will re-render the photograph in an artistic way that expresses the given concept. Since the new semantic model is the focus of this paper, the image generator is simplified to create only abstract images (by skipping the image composer) for all experiments.

Rendering images based on predicting word vectors instead of predicting the adjectives directly makes it more difficult for the image renderer to match a given adjective. However, the power comes in taking advantage of the learned semantic structure encoded in the vectors. DARCI can render images to convey any adjective that has at least some semantic relationship with any of the 145 explicitly trained adjectives. Even non-adjectives, such as 'war', can be rendered this way and essentially get interpreted as an adjective (i.e., 'war-like').

Evaluation and Results

We start with evaluating how well the semantic modeling component learns to predict word vectors from images. We then use clustering techniques to determine how well the images that DARCI produces actually reflect their intended adjective. Finally, we evaluate how clusters of images relate to each other and to the word vectors on which they are based.

Semantic Model Evaluation

We consider two metrics: *coverage* and *ranking loss*. For each adjective, the model ranks each test image by the similarity score obtained from the visual semantic model (Eq. 1). Images labeled with the adjective (positive images) should be ranked higher than images that are negatively examples of the adjective. Coverage represents how far to go down the list of ranked images in order to cover all positive images (normalized between 0 and 1). Ranking loss represents the percentage of negative images that are ranked higher than positive images. These metrics are averaged across all 145 adjectives. We compare our visual semantic model (Vector) with a binary relevance model (Binary) using 10-fold cross validation. The results can be seen in Table 1.

As expected, our visual semantic model performs worse than binary relevance on the 145 adjectives. However, the benefit is that our new model can rank images based on adjectives on which it was never trained. We chose 10 additional adjectives for which DARCI had not been trained and created a hold-out set of test images for them. We evaluated how well the model ranked images based on these new ad-

	Cross Validation				Zero-shot	
	Random	*Binary*	*Vector*		*Random*	*Vector*
Coverage	0.768	**0.533**	0.628		0.709	**0.444**
Ranking Loss	0.502	**0.297**	0.357		0.502	**0.199**

Table 1: The 10-fold cross validation image ranking results of learning the 145 adjectives (lower scores are better). We compare our visual semantic model (Vector) with a binary relevance model (Binary) that learns the adjectives directly. The binary method performs better on the 145 adjectives. However, the vector method allows the system to rank images based on adjectives it has never been trained on (Zero-shot), which we test using a hold-out set for 10 adjectives the model has never seen.

jectives (Zero-shot in Table 1). The results show that the visual semantic model is successful (i.e., better than random) at ranking the test images for the 10 new adjectives.

Image Evaluation

Evaluating how well an image conveys an adjective is a subjective task, especially for a system that is also trying to generate novel images. Usually, a human survey is necessary to arrive at a general consensus in measuring the semantic quality of images, but even then such a consensus is not always possible (or desirable).

Clustering techniques have been developed for evaluating how well images convey semantic relationships (Heath, Norton, and Ventura 2014). The idea is that images should cluster in ways that reflect the semantic similarity of the adjectives on which they were based. For example, 'scary' and 'creepy' images should cluster together more closely (i.e., be harder to tell apart) than 'cold' and 'happy' images because 'scary' and 'creepy' are more similar in meaning than 'cold' and 'happy'. By using clustering, we may not be able to objectively tell if a *specific* image conveys a *particular* adjective, but we can objectively see how well the system in general is creating images that reflect the semantic relationships learned by the vector space model. Heath et al. showed that their clustering methods were consistent with human evaluators.

Let *SEEN* refer to the set of 145 adjectives that DARCI was trained on and let *UNSEEN* refer to adjectives not of those 145. We selected two sets of 5 adjectives from SEEN. The first set consisted of semantically *similar* adjectives, while the second set consisted of semantically *distinct* adjectives. We had DARCI render 10 separate images for each adjective using the abstract rendering method (i.e., no source image). The two sets of 5 adjectives are listed and example images for each can be viewed in Figure 3.

The 51 global and local features from the visual semantic model were extracted from each rendered image. We used the EM (Expectation Maximization) algorithm found in WEKA (Hall et al. 2009) to cluster each set's collection of images (using the extracted features). We then applied two metrics, average *entropy* and average *purity*, to evaluate the quality of the clusters. The results can be seen in Table 2.

The results verify that the images for the similar set of

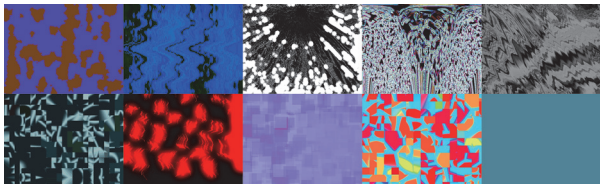


Figure 3: Example abstract images created for the adjectives referenced in Table 2. The top row (from left to right) corresponds to the semantically similar adjectives 'creepy', 'ghastly', 'scary', 'strange', and 'weird'. The bottom row corresponds to the distinct adjectives 'cold', 'fiery', 'peaceful', 'vibrant', and 'wet'.

	Similar	Distinct
Entropy	0.857	**0.714**
Purity	0.360	**0.440**

Table 2: The cluster entropy and purity results from clustering images of semantically similar adjectives compared to clustering images of semantically distinct adjectives (the adjectives are listed in Figure 3). Lower entropy is better, while higher purity is better. These results confirm that it is harder to cluster the images of similar adjectives than it is to cluster the images of distinct adjectives.

adjectives are harder to correctly cluster than are images for the distinct set of adjectives. This is evidence that DARCI is successful at rendering images that convey the meaning of adjectives relative to each other. In this paper, we especially want to focus on how well DARCI can render images based on UNSEEN adjectives, or any word for which it has never seen images.

We chose 10 UNSEEN adjectives and had DARCI render 10 separate abstract images for each. We also chose five non-adjectives and again had DARCI render 10 abstract images for each. The words are listed and example images for each are shown in Figure 4. We again used clustering to evaluate the the semantic quality of the rendered images. For each of the 10 UNSEEN adjectives, we took a SEEN adjective that was semantically similar and one that was dissimilar and had DARCI generate 10 images for each of them. We then clustered the images for the UNSEEN adjective and the images for the SEEN *similar* adjective, while separately clustering the UNSEEN images and the SEEN *dissimilar* images. Finally, we averaged the metrics of the 10 UNSEEN adjectives. We repeated this process for the five non-adjectives and the results are shown in Table 3.

The similar images are harder to cluster than the dissimilar ones for both UNSEEN adjectives and non-adjectives. This indicates that DARCI is successfully rendering the images to convey the intended words relative to each other, even though DARCI has never seen any example images of the words. DARCI is able to use the semantic structure learned from the vector space model to interpolate, or more colloquially "imagine", what images of these unseen adjectives could look like. The clustering results give us a measurable indication of DARCI's ability to render images con-

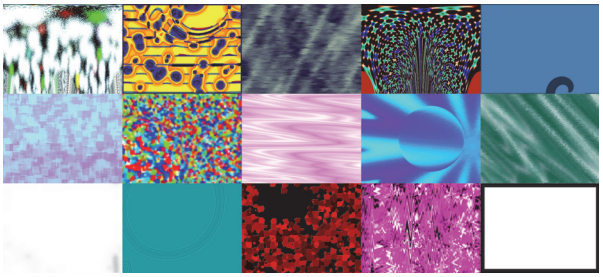


Figure 4: Example abstract images created for adjectives DARCI was never trained on and that correspond to the results in Table 3. The images of the first two rows from left to right convey the adjectives ‘bizarre’, ‘brilliant’, ‘freezing’, ‘frightening’, ‘frigid’, ‘hazy’, ‘lively’, ‘lovely’, ‘luminous’, and ‘somber’. The images of the third row convey the non-adjectives ‘Alaska’, ‘crying’, ‘fear’, ‘love’, and ‘winter’.

	Adjectives			Non-adjectives	
	Similar	*Dissimilar*		*Similar*	*Dissimilar*
Entropy	0.691	**0.480**		0.828	**0.479**
Purity	0.775	**0.850**		0.680	**0.840**

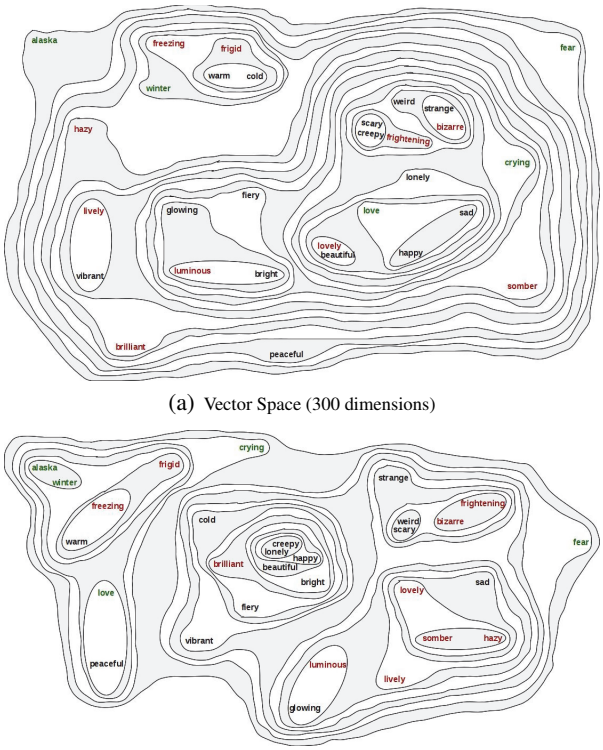
Table 3: The average cluster entropy and purity results from clustering images conveying adjectives (and non-adjectives) on which the system was never trained. The adjectives and non-adjectives used are listed in Figure 4. Lower entropy is better, while higher purity is better. The results show that it is harder to cluster images from semantically similar words than images from dissimilar words. This is evidence that DARCI is successfully rendering images that convey the intended word, even when it has never seen an example image of that word before.

sistent with the semantic structure of the words for which they were rendered.

Image Cluster Visualization

We can also visualize how the clusters of images relate to one another and to the semantic vectors on which they are based. We created a 2D visualization of how the words cluster in *vector space* and compared it to a 2D visualization of how their respective images cluster in *image feature space*. We created the 2D visualizations by using agglomerative clustering combined with multi-dimensional scaling (Pich 2009). We took the 10 UNSEEN adjectives and 5 non-adjectives from the previous experiment and chose an additional 15 SEEN adjectives that had variable semantic similarity to the 15 UNSEEN words.

For the visualization in vector space, we used multi-dimensional scaling to find an approximate 2D plot (from 300 dimensions) of the distances between each word’s vector. We then did agglomerative clustering (using EM) with the 30 word vectors and drew the resulting clusters on the 2D plot. For the visualization in image feature space, we calculated the average feature vector of the 10 separately rendered images for each of the 30 words. We then performed multi-dimensional scaling (from 51 dimensions) and agglomerative clustering in the same way we did with the word vectors. Both visualizations can be seen in Figure 5.



(a) Vector Space (300 dimensions)

(b) Image Feature Space (51 dimensions)

Figure 5: A 2D visualization of the spacial relationships between the word vectors (a), compared to the spatial relationships of their respective images (b). Red words are adjectives on which DARCI was never trained, while green words are non-adjectives. The image clusters/positions roughly correspond to the word clusters/positions. This demonstrates that DARCI was able to render images that at least partially convey the meaning of adjectives, and even of words on which DARCI was never trained, including non-adjectives.

In vector space, note the distinct clusters of similar words. Also note that the non-adjectives are generally more distant from the larger groups of adjectives, likely due to their having closer similarities to some other non-adjectives. Overall, the image clusters roughly correspond to the word clusters. In both visualizations there exist relative groupings for scary type words/images, and groupings for temperature type words/images. Even in the clusters that don’t match exactly, the relative positions of most words are similar. For example, ‘bright’, ‘luminous’ and ‘glowing’ are still generally near each other, even though they were absorbed into different neighboring clusters. Differences between the word clusters and the image clusters are to be expected as the visual semantic model learns from noisy data and multi-dimensional scaling has to approximate 2D positions from a high dimensional space. Also keep in mind that DARCI, while trying to convey the adjective in the image, is also trying to innovate and create novel images. For example,

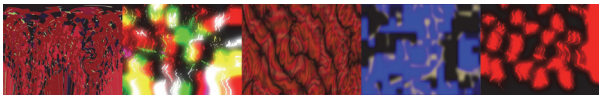


Figure 6: Five of the 10 abstract images rendered for the adjective 'fiery'. Notice the variation between different renderings as DARCI is trying to innovate, in addition to conveying the adjective.

Figure 7: Five of the 10 abstract images rendered for the adjective 'cold'. Notice that some of the images could easily be confused with 'warm' due to 'cold' being semantically related to 'warm'.

Figure 6 shows variations across different renderings of the same adjective ('fiery').

It should be noted that visual differences between words don't always correspond to their semantic differences. For example, we would expect the adjectives 'warm' and 'cold' to have distinct visual qualities. However, in Figure 5(a) we see that 'warm' and 'cold' are semantically similar and so DARCI's renderings of these adjectives can look similar. Figure 7 shows five of the 10 rendered images for 'cold'. Notice that a few of them could easily be confused with a 'warm' image. This seems unfortunate, but it is actually another indication that DARCI is accurately generating images according to the semantic relationships learned by the VSM.

Conclusions and Future Work

We have introduced a sophisticated semantic model into DARCI that enables it to create images that convey a wide variety of concepts. We have shown that the similarity of the resulting images correspond to the semantic similarity of the concepts on which they were based, which is evidence that the images do reflect their intended adjective. We have also shown that DARCI can render adjectives (and even non-adjectives) that it has never seen example images of. This ability is a rudimentary form of imagination and is analogous to a person being able to imagine, say, what a 'majestic' image might look like when told that 'majestic' is similar to 'powerful' and 'beautiful', even though the person may have never experienced the word 'majestic' before.

This simple form of imagination is not limited to images and could be applied to practically any domain. For example, a system could generate music based on the same word vectors (e.g., compose a 'happy' song), and could then produce new music to match previously unheard concepts. The VSM could even act as a bridge between different domains. A system could listen to a 'sad' song, which would be mapped near the 'sad' vector, and the system could then "imagine" a sad-like image inspired by the song.

Using a VSM for these types of creative learning problems allows for more freedom, more autonomy, and demonstrates a more robust form of intelligence. In classical machine learning, models are typically rigidly confined to the concept(s) explained by available training data, performing poorly outside this scope. In contrast, the semantic model presented here attempts a form of transfer learning from written text to image understanding/generation, which gives our system a chance to perform reasonably even for concepts it has not explicitly learned. This flexibility is especially useful for problems in the field of computational creativity, where there may not be a "best" or "right" answer.

Figure 8: Images DARCI rendered (bottom row) after being provided a source image (top row) and a concept. From left to right, the concepts are 'fiery', 'Alaska', and 'hunchback'. Although the source image was given, DARCI discovered its own way to render the image to convey the given concept.

Figure 9: Images that DARCI has rendered after being given only a concept. From left to right, the concepts are 'bizarre', 'war', 'art', 'murder' and 'hunger'.

With the success of the semantic model, we can consider the system as a whole and move beyond abstract images by having DARCI create more sophisticated art that conveys meaning for more advanced concepts. For example, a user could provide a source image and DARCI could then re-render the source image to convey any given concept. Figure 8 shows several examples of this method of rendering.

As outlined in the Image Generator Section, DARCI can also create a source image of simple icons by finding nouns semantically related to the provided concept. This collage of icons can then be artistically rendered to communicate the original concept, as shown in Figure 9. DARCI chooses to include icons based on what it has learned through the vector space model, and the result is an original image that conveys the given concept. We intend to evaluate the DARCI system as a whole to determine its creative ability to communicate meaning through visual art.

We noted that semantic differences between words don't always correspond to visual differences. One idea to overcome this is to use a hierarchical approach that locates different densities or clusters within the word vector space: a top-

level visual semantic model that learns to identify different clusters, and separate visual semantic models for each cluster that focus on distinguishing among the individual words in a given cluster.

We would eventually like to extend the ideas in this paper beyond adjectives to include nouns. We want to enable DARCI to create actual (non-abstract) pictures of nouns without relying on a provided source image or a database of icons. This will most likely require a deep learning system that leverages semantic information to discriminate between pictures of nouns, as done in other studies (Frome et al. 2013). A deep generative model could potentially generate images by visualizing how the model has learned features at various levels. Recently, deep neural systems have already had success in automatically generating images (Gregor et al. 2015; Leon A. Gatys and Bethge 2015; Denton et al. 2015).

References

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.

Coyne, B., and Sproat, R. 2001. WordsEye: An automatic text-to-scene conversion system. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, 487–496. New York, NY, USA: ACM.

Csíkzentmihályi, M., and Robinson, R. E. 1990. *The Art of Seeing*. The J. Paul Getty Trust Office of Publications.

Denhière, G., and Lemaire, B. 2004. A computational model of children's semantic memory. In *Proceedings of the 26th Conference of the Cognitive Science Society*, 297–302. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Denton, E.; Chintala, S.; Szlam, A.; and Fergus, R. 2015. Deep generative image models using a laplacian pyramid of adversarial networks. *arXiv preprint arXiv:1506.05751*.

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.

Frome, A.; Corrado, G.; Shlens, J.; Bengio, S.; Dean, J.; Ranzato, M.; and Mikolov, T. 2013. DeViSE: A deep visual-semantic embedding model. In *Advances In Neural Information Processing Systems*, 2121–2129.

Gregor, K.; Danihelka, I.; Graves, A.; and Wierstra, D. 2015. DRAW: A recurrent neural network for image generation. In *Proceedings of The 32nd International Conference on Machine Learning*, 1462–1471.

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; and Witten, I. H. 2009. The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter* 11:10–18.

Heath, D.; Dennis, A.; and Ventura, D. 2015. Imagining imagination: A computational framework using associative memory models and vector space models. In *Proceedings of the 6th International Conference on Computational Creativity*, 244–251.

Heath, D.; Norton, D.; and Ventura, D. 2014. Conveying semantics through visual metaphor. *ACM Transactions on Intelligent Systems and Technology* 5(2):31:1–31:17.

Joshi, D.; Wang, J. Z.; and Li, J. 2006. The story picturing engine—a system for automatic text illustration. *ACM Transactions on Multimedia Computing, Communications, and Applications* 2(1):68–89.

Landauer, T., and Dumais, S. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition induction and representation of knowledge. *Psychological Review* 104(2):211–240.

Leon A. Gatys, A. S. E., and Bethge, M. 2015. A neural algorithm of artistic style. *Computing Research Repository*.

Liu, Y.; Zhang, D.; Lu, G.; and Ma, W.-Y. 2007. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition* 40(1):262–282.

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

Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. 2013. Efficient estimation of word representations in vector space. In *Proceedings of the International Conference on Learning Representations*.

Norton, D.; Heath, D.; and Ventura, D. 2015. Accounting for bias in the evaluation of creative computational systems: An assessment of DARCI. In *Proceedings of the 6th International Conference on Computational Creativity*, 31–38.

Norton, D.; Heath, D.; and Ventura, D. 2016. Annotating images with emotional adjectives using features that summarize local interest points. *IEEE Transactions on Affective Computing, Under Review*.

Pich, C. 2009. *Applications of Multidimensional Scaling to Graph Drawing*. Ph.D. Dissertation, University of Konstanz.

Rapp, R. 2003. Word sense discovery based on sense descriptor dissimilarity. In *Proceedings of the Ninth Machine Translation Summit*, 315–322.

Salton, G. 1971. *The SMART Retrieval System—Experiments in Automatic Document Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

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

Turney, P. D. 2006. Similarity of semantic relations. *Computational Linguistics* 32(3):379–416.

Wang, F. 2011. A survey on automatic image annotation and trends of the new age. *Procedia Engineering* 23(0):434 – 438.

Zhu, X.; Goldberg, A. B.; Eldawy, M.; Dyer, C. R.; and Strock, B. 2007. A text-to-picture synthesis system for augmenting communication. In *Proceedings of the 22nd National Conference on Artificial Intelligence - Volume 2*, 1590–1595. AAAI Press.