# CONCEPTUAL BLENDING FOR THE VISUAL DOMAIN

Alexa Steinbrück

<alexasteinbrueck@gmail.com>

10009728

Bachelor Thesis
Credits: 18 EC

Bachelor Opleiding Kunstmatige Intelligentie

Faculty of Science
University of Amsterdam
Science Park 904
1098 XH Amsterdam

*Supervised by:*

**dr. F.M. Nack**
Informatics Institute
Faculty of Science
Universiteit van Amsterdam
Science Park 107
1098 XG Amsterdam

June 28th, 2013

**Abstract**

*Conceptual blending* has been presented as a model for the creative cognitive process, but most theoretical work emphasizes the analysis of existing blends rather than the concrete mechanisms for the construction of novel blends. In consequence, artificial creative systems that implement conceptual blending do not provide a formalization of all steps of the process. *Conceptual blending for the visual domain* (CBVD) is a framework that formalizes the entire process of conceptual blending while applying it to the visual domain. Since this is a work in progress, we describe the first preliminary results and try to provide the reader with a clear image of the framework to be followed.

# Contents

# 1 INTRODUCTION

Creativity is an exciting feature of human intelligence. While it is often regarded as a fascinating and mysterious quality, it is the task of computational creativity to understand and formalize it as far as is possible. As a key researcher in the field of creativity, Margaret Boden took a great step forward in demystifing the creative process. Boden [Boden, 1991] describes three types of creativity: combinatorial, exploratory and transformational. The use of the combinatorial type of creativity generates a novel combination of preexisting ideas, states, concepts or objects.
A great deal of human creativity can be understood as belonging to the combinatorial type. A prominent example within the fine arts is Duchamp's *Fountain* from the year 1917, which combines a familiar object (the toilet) with an unfamiliar context (the art exhibition). The surrealist painter Magritte masterly combined different concepts to synthesize new concepts which are rich in analogy and meaning (e.g. *Le viol*, 1935).
The combinatorial perspective is particularly interesting for Artificial Intelligence, because it allows creativity to be modeled as a search process through the space of possible combinations, in order to find the best combination possible. Gilles Fauconnier and Mark Turner recently developed the theory of *Conceptual Blending* [Fauconnier and Turner, 2001], which describes the process of combinatorial thinking. Conceptual blending involves two input concepts that, according to a given structure mapping, will generate a novel concept, called blend. This new concept will maintain some structure from the input concepts, but sometimes also exhibit new structural features that emerge through the process. Conceptual blending has been described as a fundamental process of not only creativity, but also general every day thinking. The importance of this theory for AI can therefore not be valued too highly. However, conceptual blending is not yet a popular topic within AI research.
The conceptual blending theory also provides various interesting starting points for a formalization within the field of computational creativity. However, concrete attempts are rather sparse and those that do exist do not formalize the whole process of conceptual blending.
The here presented project, *Conceptual Blending For The Visual Domain* (CBVD), is motivated by the goal to formalize the entire process of conceptual blending and apply it to the visual domain. The goal is to build a system that is able to generate images that can be rated as creative. The research question focuses on how appropriate knowledge about an image can be gathered and how concepts can be semantically related to each other in order to make a creative combination of them, that is not trivial.

The architecture of CBVD consists of 5 distinct modules: The first two modules are dedicated to knowledge acquisition and enable a dynamic construction of the knowledge base.
*Module 1* detects visual features (e.g. circular subparts) in an image by the use of different algorithms from Computer Vision. *Module 2* gathers semantic knowledge about the concept depicted in an image with the help of common sense ontologies. The remaining modules deal with the combinatorial process. *Module 3* implements the rules for the selection of the best suited pair for combination. *Module 4* selects the parts of each image that should be involved in the blending. Finally, *Module 5* performs the combinatorial process directly on images on a pixel level and generates the final output image. CBVD has been implemented in the programming language python.
CBVD aims to be a framework, therefore it can be considered as an invitation for extension. In consequence, the architecture is built in a modular way to support the addition of new features

and methods. This paper explains the architecture of CBVD. First, we review relevant literature that covers related work in AI and theoretical issues from creativity research. Next we examine the theory of conceptual blending in detail and explain how this theory is concretely implemented in our system. Subsequently, we discuss the organization of CBVD, explain what each module does in detail and how it has actually been implemented. We then illustrate CBVD's operations on several examples, followed by a detailed evaluation of the system's output. Finally, the last section is dedicated to the discussion of our approach and provides a perspective on future work.

# 2 LITERATURE REVIEW

The literature studied covers theoretical aspects of (computational) creativity as well as related work.

Finke [Finke et al., 1992] developed a psychological model of creativity called the *Geneplore Model*. Under this model, the creative process is divided in two distinct phases. The first phase is the generative phase, where the emphasize is on quantity. The generative process includes knowledge retrieval, idea associations, synthesis, transformation and analogical transfer. The second phase is the explorative phase, where the ideas generated in the first phase are expanded.

The structure mapping engine [Falkenhainer et al., 1989] is an implementation of Gentner's psychological theory of structure mapping [Gentner, 1983]. This theory describes how analogies are constructed by the mapping of structural features. Gentner's core idea is that an analogy is a mapping from one domain of knowledge into another one (from a base domain to a target domain). The engine provides a toolkit with matching algorithms and can be used both as a cognitive simulation of human analogical processing and as a component in a larger machine learning system. The engine has been used in numerous projects and its popularity may lay in the fact that it ignores surface features and finds matches between semantically very different things if they have a similar representational structure.

Margaret Boden aims to frame creativity into computational paradigms. Therefore she has been a major inspiration for research that aims to implement creativity in artificial agents. Considering combinatorial creativity, Boden rejects the idea that the combination of ideas is a random process, and emphasizes the importance of an "intelligible, though previously unnoticed link between them, which we value because it is interesting - illuminating, thought provoking, humourous" [Boden, 1991]. Boden also discusses the role of unpredictability (surprise) and intuition.

While systems of computational creativity have been around for several decades now, most of them work within the domain of language. Poetry generators, story engines, joke generators - implementations of these types are numerous. In contrast, systems that apply creativity to the visual domain are rare. One of the first was AARON (as described by P. McCorduck [McCorduck, 1990]), a program created by Harold Cohen in 1973, that is still under development and has several different implementations: Abstract-AARON draws landscape-like motives by the use of a production system with concrete "if-then" rules which specify what decision should be taken in every given situation. These decisions may concern if a a line should be continued, and if so in what direction. Acrobat-AARON is a more elaborated variant since it is able to plan its actions before excecuting them. The generated drawings by AARON are shown in art galeries

around the world. McCorduck considered Aaron as the first profound connection between art and computer technology.
Considering the quite recent theory of conceptual blending, a few computational implementations have been proposed; the most relevant of them are discussed here briefly.
The Alloy blending engine, which is a part of the GRIOT poem generator [Goguen and Harrell, 2004], is the earliest attempt to implement the theory of conceptual blending within an artificial creative system. GRIOT generates single lined poetry in response to user input.
Another project that was directly inspired by Fauconniers theory of conceptual blending is DIVAGO [Pereira, 2005]. In the very first implementation, DIVAGO was meant to be a a pure concept generator since it generates new concepts out of previous knowledge. Subsequent research conducted by Pereira (2006) involved an implementation of the blending theory for the linguistic and the visual domain. The latter includes the creation of mythical creatures with 3-D graphical models. Two things are noteworthy about the implementation of DIVAGO: Firstly, the creative process is driven by a goal (e.g. "Create something that can fly and is a transport means"); and secondly, the inputspaces are not selected by DIVAGO itself, but rather given to it. The system starts with a pair of concepts (e.g. horse and bird) and tries to create possible blends that meet the prespecified goal. The blending is done by selecting subparts or properties of one concept and combining it with subparts and properties of the other. The search space is exponentially growing since every combination of parts is possible. Therefore DIVAGO uses a genetic algorithm to stochastically sample the space of possible blends.
A precursor of DIVAGO is BLENDER [Pereira and Cardoso, 2002]. Blender produces visual blends in a logo like visual language similar to line drawings. The knowledge base of Blender is manually constructed and considers only superficial knowledge. Blender was tested on the task to combine the concepts "boat" and "house", where all possible blends are generated in an exhaustive way. The experiment was motivated by the need to show what a framework like BLENDER can actually achieve, instead of demonstrating its value as a finished work.

# 3 THEORY OF CONCEPTUAL BLENDING

This section reviews the theory of conceptual blending as described by Fauconnier and Turner (2001). The theory of conceptual blending relies on the notion of *mental spaces*. A mental space contains a number of concepts and their inter-connections. Mental spaces can be dynamically constructed during a discourse (e.g. conversation) but they can also preexist in the knowledge of the agent.

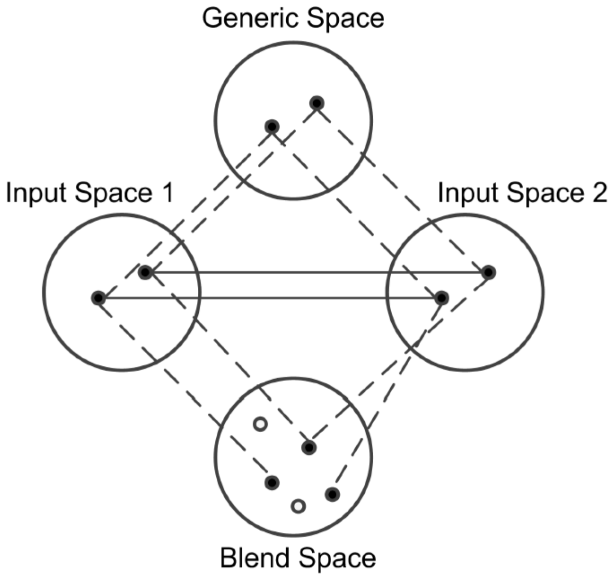


Figure 1: The four-space model of Conceptual Blending

Consider the sentence “This surgeon is a butcher”, in which we can detect the blend of two mental spaces: The mental space for “surgeon” includes the surgeon and other relevant concepts such as his scalpel, the patient, and the surgery room. The mental space for “butcher” includes among other things: the butcher, the knife, the abatroir. Conceptual blending needs two or more mental spaces as input (*input spaces*). These mental spaces must have an analogy in structure to enable a mapping of parts from one space to the other. In the surgeon-as-a-butcher example this mapping works as follows: The knife of the butcher is mapped to the scalpel of the surgeon, the dead animal is mapped to the patient, and the abattoir is mapped to the surgery room. The selected elements form the input spaces are then projected into the *generic space*. Finally a *blend space* is constructed by further elaboration on the generic space, e.g. by pattern completion. The four-space model is represented in *Figure 1*. Black dots denote elements in the mental spaces, solid lines are the mappings between these elements and dashed lines represent the correspondences between all these elements throughout the four spaces. The hollow dots denote emergent structures in the blend space.

Li et al. [Li et al., 2012] demonstrated that there are two types of creative artifacts that can be produced by conceptual blending: Blends that have a use in communication (as the surgeon-as-butcher sentence), and blends that represent selfstanding novel concepts. These selfstanding novel concepts are ubiquitous in fiction: Dragons, for example, are blends of the concepts snake, cat and bird of prey. Many other mythical creatures are blends of animals that exist in reality. But blending is not limited on creature generation: In *Star wars*, a "lightsaber" represents a blend of the concepts “sword” and a “laser emitter”. And the popular Japanese manga “Doraemon” (Fujio 1974-1996) features gadgets such as a telephone that transmits flu instead of sound.

Fauconnier and Turner describe the blending process with 5 steps:

- Selection of the (two) inputspaces

- Selection of elements for projection
- Creation of a generic space/blend space
- Elaboration

However, Li et al (2012) pointed out that that the blending theory does not explain for all 5 steps how they are excecuted in detail. Firstly, it has not been specified according to which criteria the two inputs spaces are selected. A full computational implementation of the blending theory should select input spaces by itself, rather than assume them as given (such as in DIVAGO).
Secondly, no detailed procedures have been described for how the elements from the input spaces could be chosen for projection. A system that performs a projection of all available elements would risk a combinatorial explosion.
Thirdly, blend elaboration should have a stopping criteria, in order to keep it within the boundaries of efficiency.

# 4 THE FORMALIZATION OF CONCEPTUAL BLENDING IN OUR SYSTEM

Our system aims to be a complete formalization of conceptual blending. We will now explain how we applied conceptual blending to the visual domain and explain our proposal to fill the above mentioned gaps in concrete explanation of the process.
*Figure 2* illustrates the correspondences between the modules in our system and the steps within the process of conceptual blending. Before we can talk about the five steps, we have to explain what we consider a *mental space* within our system.
Instead of defining a mental space in semantic terms (such as with the doctor/butcher example, where the elements of the mental space where objects such as “knife” etc), we define it on a purely visual basis. That means that a mental space has elements that describe visual features that can be recognized in a visual representation of it, such as a photography. The mental space “traffic light” will contain elements such as: a green circular shape, a yellow circular shape, a red circular shape and a black rectangular shape that surrounds them. In addition we can add information about how these parts are structurally related to each other: Considering the three circular shapes, we can add the information that they are distributed regularly on a vertical line, to the mental space. Adding structural information like this defines the mental space more precisely and can help to make stronger analogies to other mental spaces. The construction of the mental spaces is implemented in *Module 1*.

| Module in CBVD | Content | Relevance in the process of conceptual blending |
|---|---|---|
| **Module 1** | Gathering visual characteristics | Creating the mental spaces |
| **Module 2** | Gathering semantic distances | Gathering knowledge necessary for the selection of input spaces |
| **Module 3** | Choosing the best combinations of images | Selecting the inputspaces |
| **Module 4** | Selecting the image parts to be blended | Selecting the elements for projection |
| **Module 5** | Generate the image | Creation of the blend space & Elaboration |

Figure 2: Correspondences between the the 5 steps of Conceptual Blending and the modules in our system

Now we will discuss how we interpreted the actual steps of the blending process.
The first step has been defined as the selection of the input spaces. We already mentioned that this step has not been sufficiently explained in the theory of conceptual blending. With our system we propose a concrete definition of how this process works and on which criteria it relies. Within our framework the definition of a good combination is 2-fold: A good combination is when two input spaces share the same visual features and exhibit a big semantic distance. Since the mental spaces have only elements that contain information about visual features, we can say that we select the input spaces according to the amount of visual analogy between them. Visual analogy is defined as having same geometric properties, e.g. the existence of circular shapes or color properties, the number of parts and structural information (how they are distributed). The first criterion (same visual features) ensures an "effortless" blending at the visual level. The second criterion (semantic distance) ensures the surprising effect, that is connected to the criterion of *novelty*, which is an essential property of any creative production: Concepts that are normally not directly associated are "effortlessly" combined in the generated image. To put it in terms of Poincaré [Poincare, 1982]: "Among chosen combinations the most fertile will often be those formed of elements drawn from domains which are far apart (...) Most combinations so formed would be entirely sterile; but certain among them, very rare, are the most fruitful of all." The knowledge necessary for fulfilling the second criterion is gathered in *Module 2*. While the integration of both criteria in order to select the input spaces in done in *Module 3*.

The second step within the process of conceptual blending is the selection of elements from each inputspace for projection. In our system we define the notion of *base image* and *substitute image*. A base image is the inputspace which contributes its general structure (the environment that surround the parts). The substitute image is the inputspace, which is "pick to pieces" and provides the elements that are blended into the base image. Therefore, the projection of one element to the other is implemented as pure substitution of an element in the base image with the corresponding element in the substitute image. The selection of the parts is implemented in *Module 4*.

The third step of the process is the creation of a *generic/blend space*. This is done by a simple pasting of the corresponding elements from substitute image in to the base image.
Finally, the fourth step of elaboration is implemented in our system as post-processing on the color level in order to make a visually smooth integration. Step 3 and 4 have been implemented in our *Module 5*.

# 5 ARCHITECTURE

## 5.1 The image base

Within the visual domain, several types of visual representation are possible: Simple geometric shapes (less able to represent real world concepts), line drawings, drawings that include lines as well as filled shapes, illustrations of single concepts (like in an encyclopedia), grayscale photographies and color photographies. We chosed to focus on the latter. Therefore our image base consists of photographies of real world motives. Some constraints are set on the choice of images for the image base: Firstly, the images have to depict a certain concept clearly in the foreground. Secondly the depicted concept has to be well known in common sense, that means that the name of the concept has to have an entry in the common sense ontology *Conceptnet*. Our image base consists of 15 images that depict the following concepts: Apple tree, bicycle, bowling, eye, globe, lilly pond, man, medicament, pearl chain, pizza, police, snail, tower, traffic lights, universe. Each image is labeled by the name of the concept that it depicts.

## 5.2 Module 1 - Gathering visual characteristics

### Defining the concept area

The first step in understanding an image is to define where the concept is exactly depicted in the image. This is a well known task in computer vision called "image segmentation". The core idea is that the pixels of an image can be labeled as belonging to a certain concept or not. A subtask of this task is to detect the foreground of an image. Since our database consists only of images that depict a single concept clearly in the foreground, we can work with the hypothesis that the detected foreground area will be more or less equal to the area which depicts the concept. Foreground detection is much easier than general image segmentation because it involves less machine learning. The *Grabcut* algorithm [Rother et al., 2004] is an efficient method for foreground segmentation in still images, that makes use of both color and edge information. The exactness of this first step is crucial for the quality of the results, because a too large boundary could result in the selection of image parts that do not belong to the concept itself. Therefore we propose an interactive variant of the *Module 1*, in which a bounding rectangle around the outer points of the concept area can be defined in order to enhance the performance of the Grabcut algorithm (see *figure 3*).

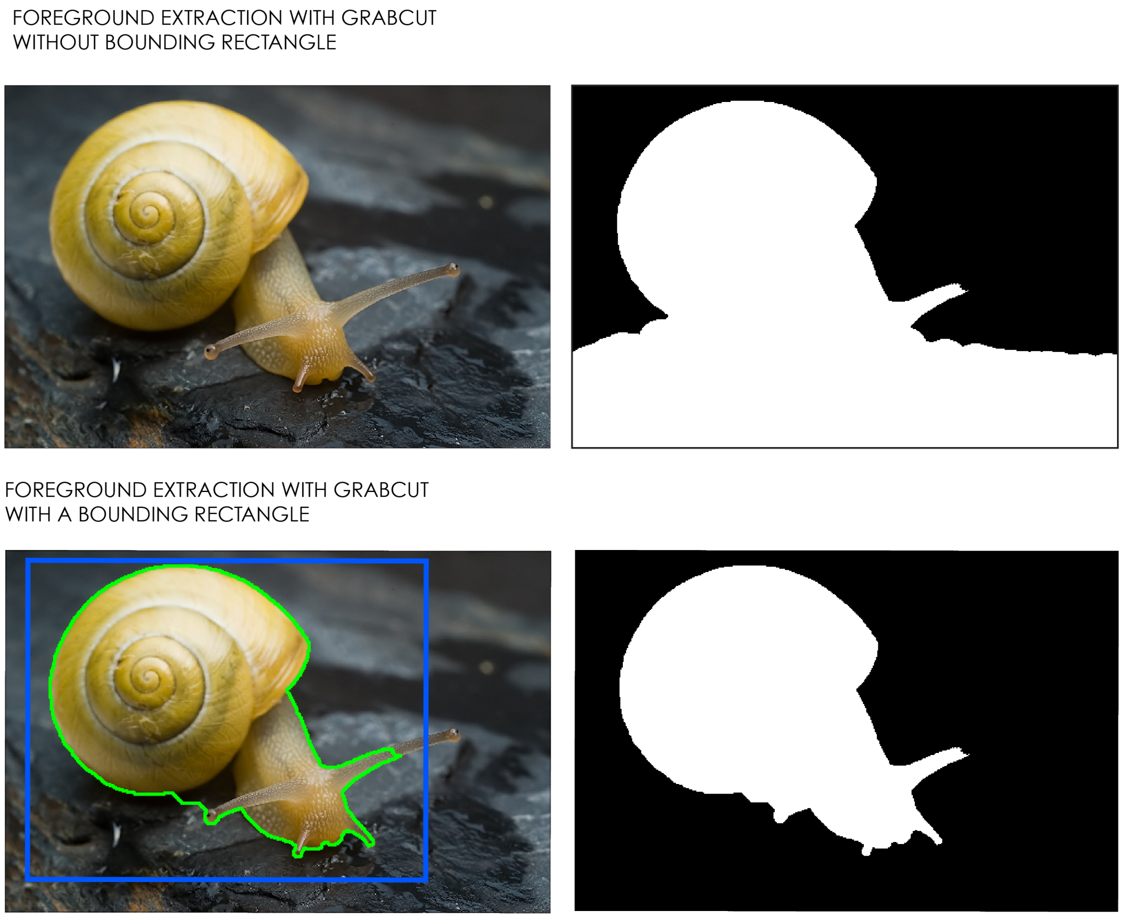


Figure 3: Defining the concept area with *Grabcut*

**Part detection**

Once the boundaries of the depicted concept are known, we can concentrate on gathering knowledge about the visual features of the concept itself. What we are interested in, is to detect if the concept has subparts and to define their contour. The existence of subparts makes conceptual blending possible, since these are the elements that are blended. The fact of detecting no subparts in a concept would disqualify the concept as a candidate for a blend. We look for subparts that have a characteristic geometric shape, such as a circle. Our approach for circular partdetection is twofold:

First an image is split into its channels (red, green, blue). Then each channel is binarized by applying a low threshold. This will result in three gray scale images. The next step is to reduce noise by the two morphological processes *dilation* and *erosion*. These processes respectively enlarge and subsequently diminish white areas in a binary image and are also useful when the image consists of overlapping or adjacent parts that should be detected as discrete ones.

The next step is to detect edges in the image by using *Canny edge* detection. Consequently the found edges from all three channels are combined in one image. The next step is to detect contours in the image that enclose a subpart.

Each detected sub part contour is then tested on being circular. In statistical terms, circularity can be calculated as the (normalized) standard deviation of the distances of the contour points to the center of the contour. After a threshold was defined, every contour can be classified as a circle or a non circle. If no circles can be detected, the algorithm starts the next iteration and increases the threshold for binarization. The intuition behind increasing the threshold is that a part has to have a low variety of colors by definition. Increasing the threshold is therefore approaching the respective color of the part in order to distinguish it from the rest.
However, this intuition is based on a hypothesis. It might be true for the example of the pizza, where the circular parts are almost monochrome. But the hypothesis fails for the example of the globe (see *figure 5b*), where the circular parts exhibit a strong (color) variance. For this type of images, thresholding is not the appropriate approach. More promising is the following: Applying canny edge detection to the gray scaled image without thresholding it, and then finding contours and check them for circularity. This approach is taken when the first approach does not produce any results.

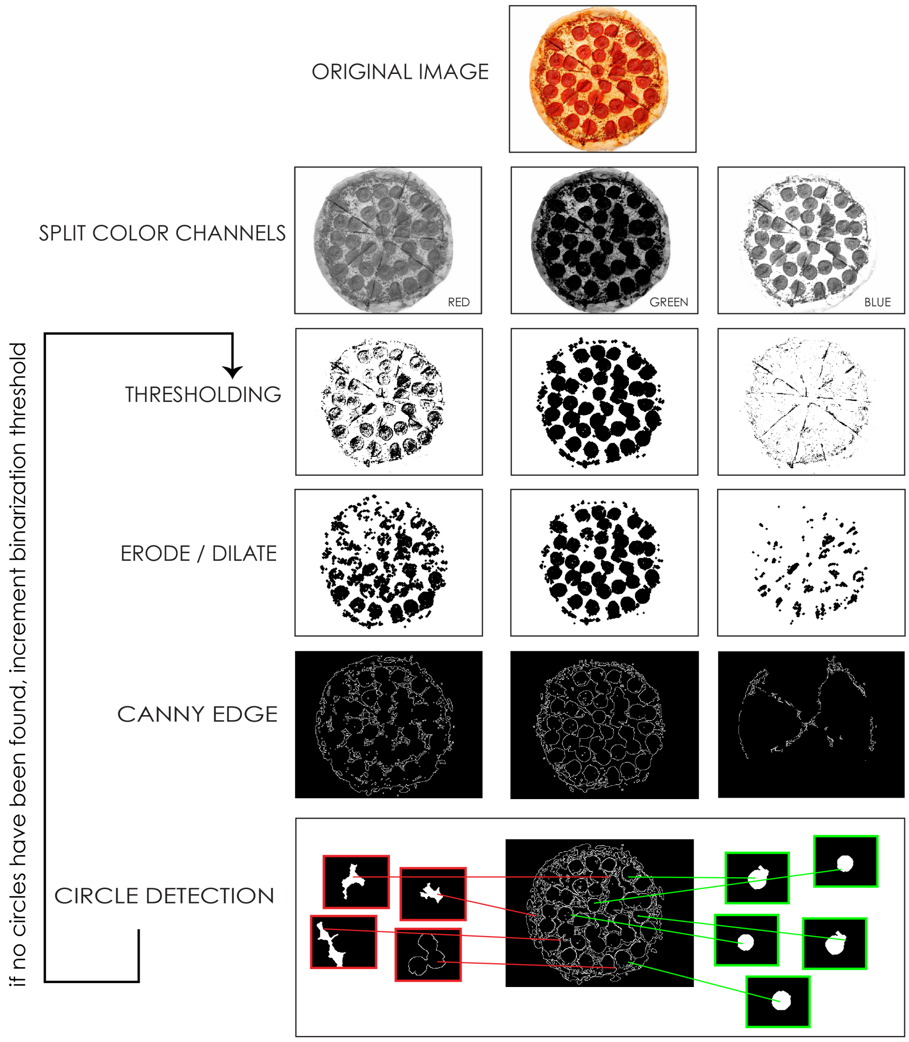


Figure 4: An iterative approach for part detection

However, the here described algorithm didn't give the expected results all the time. Especially for images where the the color variance within the part is very high (*Figure 5 b*), or where the circular parts are almost monochrome and overlapping (*Figure 5 a*) the algorithm does not detect a single area where the part should be. An algorithm that successfully solves this problem should probably be more sensible to local contrasts.
Our emphasis within this project was on building a framework for applying conceptual blending to the visual domain, and not on computer vision; therefore we decided to define the part areas our self, since a good performance of the part detection module is crucial for the subsequent process.

(a) Overlapping circles in an image (b) Circular subpart with a lot of variety

Figure 5: Detecting circular parts

**Additional visual characteristics**

Apart from the detection of the geometric shape of a subpart, we can assess the amount of *visual information* that this part contains. The measurement of visual information content is based on the amount of color variation and the numbers of edges. We use the normalized average of the two values as a measurement for recognizability. Normalization occurs by the size of the part, in order to fairly compare the visual information of parts that have different size.

## 5.3 Module 2 - Gathering semantic knowledge

Knowledge about the semantics of a concept can be acquired using ontologies like Wordnet [Miller, 1995] and Conceptnet [Speer and Alonso, 2007]. What we are interested to extract from these ontologies is knowledge about how certain concepts relate to each other, since the idea is to combine concepts that are not easily associated. A common measure of associativity between concepts is semantic similarity. Semantic similarity is a value that is assigned to a set of words/concepts based on the likeness of their semantic content. Ontologies such as Wordnet and Conceptnet offer a graph of related concepts from which semantic similarity can be extracted as the length of the shortest path between two concepts. However, this length can differ greatly between different ontologies and these differences express the different interpretations

of semantic similarity.
We need to find a measure that captures the human intuition of semantic distance. Wordnet is a strongly hierarchical semantic network that emphasises the hyponymy (is-a relation) and meronomy (has-a relation). Conceptnet in contrast emphasises less taxonomic relations, but offers a wide range of other relations that express more structural connections between concepts, such as: usedFor, atLocation, CapableOf, MadeOf, Causes, Desires, LocatedNear, HasProperty, etc. Considering the example of the concepts “wedding” and “bride”, the two ontologies would give very different results. The two concepts are semantically close but structurally unlike. In a taxonomical ontology such as Wordnet, these two concepts would be far apart, resulting in a low value for semantic similarity. In contrast, within Conceptnet the semantic similarity would be high, since the concepts can be directly connected through a "atLocation" relation. The Conceptnet ontology models the way how humans associate far better than Wordnet, since humans relate concepts not only in a hierarchical but often in a more practical way. Therefore a distance measure based on the relations in conceptnet rather than wordnet seems the right choice.
Conceptnet proposes two ways for accessing the corpus. One way is to download the corpus and use the python interface to search through the graph. Calculating the pathdistance with this interface is very slow, since the only way to do it is by a breadth first search which expands the nodes through all types of relations.
The second way is to use the Conceptnet REST API [1] , which gives access to a bigger, since crowd sourced corpus, and various tools including a function to calculate semantic similarity. This function calculates the semantic similarity in an acceptable amount of time. However, Conceptnet neither offers a documentation on how semantic similarity is calculated, nor what the actual path between the concept looks like (all intermediate relations and nodes/concepts). Still, the outcomes are very good and meet the human common sense intuition of semantic distance. Therefore we decided to use the function as a "black box".

## 5.4 Module 3 - Selection of the best suited pairs

We already explained that semantic distance and similar visual features are the two key criteria for how the inputspaces are chosen. Here we want to explain the selection process in detail. The idea of selection is to prune the search space in an iterative way in order to find the best pair possible. While pruning the search space, the emphasis is on the criterion of similar visual features. The criterion of semantic distance is considered at the very end of the pruning process. This approach avoids that concepts which don’t share any visual similarities are forced into a blend, which would result in a visually odd outcome.
The first pruning of the search space is done by making subsets of images that have detected parts. As we mentioned earlier, an image in which no parts have been detected, would be not considered in the selection process, since the subsequent blending relies on the existence of parts or features of the image, that can be blended. Subsequently we continue pruning the search space by making subsets of images that share visual properties. What we focused on in our research was the property of having circular shaped parts. Having pruned the search space to these subsets, we can finally search within them for the concept pairs that show the biggest semantic distance. Having found these pairs, they have to undergo a last check which again involves the visual properties: Since we cut out parts of one concept, we deprive them of

[1] `http://csc.media.mit.edu/docs/conceptnet/webapi.html`

their context. A part that would have a meaning when surrounded by a certain context, would loose its meaning or recognizability when presented in isolation. Since we want to preserve the recognizability of each image part when we change its context by blending it with another image, we have to check the image parts for their recognizability. Recognizability of a image part is assessed by the amount of visual information that the part contains. If both images have parts that fall below a certain threshold of recognizability, the pair of images is rejected, and the search process continues.
The output of the module is a list of the 5 best suited pairs for a conceptual blending.

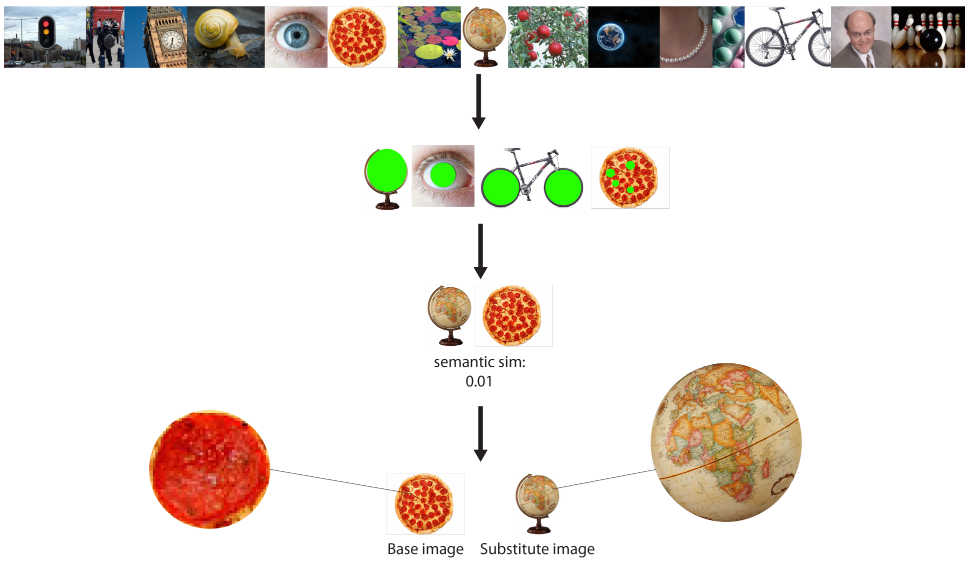


Figure 6: The selection process as implemented in *Module 3*

## 5.5 Module 4 - Selecting the image parts to be blended

Before the blending of each pair can begin, a choice has to be made regarding which image of the pair is used as a *base image* and which one as a *substitute image*. A substitute image is defined as the input image where the relevant parts are cut out and pasted on the other image, called base image. The first criterion that a substitute image has to fulfill is a high recognizability of its subparts, since these subparts are deprived of their context. The second criterion is of a more esthetical nature and regards the size of the subparts. A substitute image has to have subparts that have a bigger size than the subparts of the other image in order to avoid excessive upward scaling which would result in a visually unpleasing outcome.
Having defined the role for each image in the blending process, we have to compare the number of their subparts. Often it happens that the substitute image has less parts than the base image. In this case we duplicate the parts of the substitute image until it has the same number of parts as the base image, in order to replace all parts in the base image.

### 5.6 Module 5 - Generate the blending and further elaboration

For each part of the substitute image, a mask is generated and applied, resulting in a cutout of the original substitute image. Then the size of the part in the base image and the size of the corresponding part in the substitute image are compared and the corresponding ratio is used as a scaling factor for the masked substitute image.
After scaling the parts to the right size and calculation of the position where they should be pasted in the base image, we adapt the color of the parts of the substitute image by *tinting*. Tinting means to remove the original color of an image and recoloring it in a given hue. Therefore we detect the main color of the part in the base image which is going to be replaced, and tint the corresponding part in the substitute image in this color. The asthetic benefit of the step of tinting is that the pasted parts are better integrated in the base image, resulting in a coherent image that gives the impression that there is actually one single concept depicted instead of a mere cut and paste of parts from two images.
The last step before the composing is to apply a Gaussian blur to the mask of the substitute image, in order to avoid a hard border between the pasted parts and the baseimage. Finally each scaled and tinted part image is pasted on the base image at the right place.

## 6 RESULTS

The output of the system is a set of the 5 best rated results. The rating occurs in the selection process where the search space is pruned until the 5 best pairs of concepts are found. The resulting images depict respectively the blends of the concept pairs [snail, traffic lights], [globe, bicycle], [police, snail], [medicament, globe] and [man, eye].

Figure 7: Output 1


Figure 8: Output 2

Figure 9: Output 3

Figure 10: Output 4

Figure 11: Output 5

# 7 EVALUATION

Considering human creativity, two fundamental criteria are almost consensual: To rate something as creative, it must be somehow new and surprising (an aspect often connected to *novelty*); And it must have a certain *value* (an aspect often connected to usefulness and meaningfulness). The problem of assessing the creativity of artificial systems is currently in need of research. Graeme Ritchie [Ritchie, 2001] proposed recently a set of generic criteria that can be used to assess the creativity of almost any creative system, independent in which domain it works. Ritchie asserts that, apart from novelty and value, there is a third criterion, *typicality*, which is especially important for artificial creative systems. Typicality measures to which extent a certain result belongs to the genre that it was supposed to belong to. Ritchie argues that it is hard for artificial systems to fulfill this criteria, and mentions the examples of recent joke and poetry generators whose productions can hardly be categorized as respectively jokes or poems. As a result, he continues, the criteria of typicality should be emphasized more than novelty within the assessment of computational creativity. Before we explain his 14 criteria in detail, we have to give a set of definitions:

**B** Basic item (an entity that a program produces);

**I** The inspiring set - the set of basic items that implicitly or explicitly drove the development of the program;

**R** the set of results produced by the system;

**typ** typicality of the items;

**val** value of the items.

The 14 criteria developed by Ritchie are summarized in Figure 12.
The first 10 criteria can be seen as an elaboration on the two key criteria *value* and *typicalness*, in the sense that additional metrics are calculated out of these two. The remaining 4 criteria address the metric of novelty. Criteria 1-2 estimate the typicality of the results and criteria 3-4 their value. Criteria 5-8 consider the ratios between typical and good results. Criteria 9-10 examine the ratio of reinventions. Reinventions are results of the system that are identical with items from the inspiring set and its measurement is Ritchie's approach to address the criterium of novelty. It is important to note, that, instead of measuring novelty of a result on a continuous scale, Ritchie proposes a binary classification in "reinvention" and "non reinvention". Once the results are classified to belong to a certain class, the assessment on the two key criteria can be repeated, in order to get additional information. This is achieved by the criteria 11-14. Ritchie also introduces the three threshold variables *alpha*, *beta* and *gamma*, which are used to distinguish highly rated results. *Alpha* and *beta* belong to the metric of typicality and specify the boundaries for two classifications: A result would be classified as typical if its typicality ranges from alpha to 1, and as atypical if it ranges from 0 to beta. For the metric of value, Ritchie proposed a single threshold, *gamma*. Results that are assessed with a value under gamma, are classified as "bad results", while values above gamma will classify the result as a "good result".

| Crit. | Formalization | Informal meaning |
|---|---|---|
| 1 | $AV(typ, R) > \theta_1$ | Average typicality |
| 2 | $ratio(T_{\alpha,1}(R), R) > \theta_2$ | Ratio typical results / all results |
| 3 | $AV(val, R) > \theta_3$ | Average quality |
| 4 | $ratio(V_{\gamma,1}(R), R) > \theta_4$ | Ratio good results / all results |
| 5 | $ratio(V_{\gamma,1}(R) \cap T_{\alpha,1}(R), T_{\alpha,1}(R)) > \theta_5$ | Ratio good typical results / all results |
| 6 | $ratio(V_{\gamma,1}(R) \cap T_{0,\beta}(R), R) > \theta_6$ | Ratio good atypical results / all results |
| 7 | $ratio(V_{\gamma,1}(R) \cap T_{0,\beta}(R), T_{0,\beta}(R)) > \theta_7$ | Ratio good atypical results / atypical results |
| 8 | $ratio(V_{\gamma,1}(R) \cap T_{0,\beta}(R), V_{\gamma,1}(R) \cap T_{\alpha,1}(R)) > \theta_8$ | Ratio good atypical results / good typical results |
| 9 | $ratio(S_{\mathcal{B}}(typ, val) \cap R, S_{\mathcal{B}}(typ, val)) > \theta_9$ | Ratio results in the inspiring set / inspiring set |
| 10 | $ratio(R, S_{\mathcal{B}}(typ, val) \cap R) > \theta_{10}$ | Ratio all results / results in the inspiring set |
| 11 | $AV(typ, (R - S_{\mathcal{B}}(typ, val))) > \theta_{11}$ | Average typicality of new results |
| 12 | $AV(val, (R - S_{\mathcal{B}}(typ, val))) > \theta_{12}$ | Average quality of new results |
| 13 | $ratio(T_{\alpha,1}(R - S_{\mathcal{B}}(typ, val)), R) > \theta_{13}$ | Typical new results / new results |
| 14 | $ratio(V_{\gamma,1}(R - S_{\mathcal{B}}(typ, val)), R) > \theta_{14}$ | Good new results / result |

Figure 12: The 14 criteria developed by G. Ritchie

While the set of criteria seems as a quite demanding examination of computational creativity results, a closer look reveals plenty of room for interpretation: Firstly, Ritchie didn't define

neither how typicality nor value should be measured. Secondly the thresholds are open to be defined in accordance to the specific system. Thirdly, it is an open question how an inspiring set has to look like. Finally, the size of R is an undefined variable.
In the face of this indeterminacy, Ritchie's criteria should be better considered as a loose framework for evaluation. Nevertheless they have been applied multiple times to creative systems, and can be considered as the most accepted evaluation method until now. Usually, when Ritchie's criteria are applied to a concrete system, thresholds are defined in accordance to the specific task at hand and even the interpretation of how value or typicalness is measured is chosen with respect to the system (and its goals). In order to define the thresholds and interpretations that suit to our system, we have to consider how similar systems achieved this definition. First consider the metric of typicalness:
For the assessment of DIVAGO [Pereira, 2005] for example, the metric of typicality was substituted by its opposite novelty and calculated by *edit distance* to the inspiring set. In other words, novelty is measured by the number of cut-and-paste operations that have to be made on a item from the inspiring set, in order to achieve the final result. The decision to reject a metric of a well accepted evaluation framework and replace it with another one, seems surprising in its audacity, however, if the decision is well justified, there should be no problem. Since our system is not supposed to produce items that belong to a certain genre, the metric of typicalness is quite irrelevant. We also don't agree with Ritchie, that the criteria of novelty is so hard to achieve for a artificial creative system, that it is justified to marginalize it in an evaluation.
Instead we think that our system has a significant strength in producing novelty, and that this strength deserves to be evaluated in detail. Since our system has several similarities with DIVAGO, such as that they are both based on conceptual blending and the generating process is a series of cut and paste operations, it seems appropriate to consider their way to calculate novelty. We would then calculate novelty by the number of cut and paste operations. In our system this number depends exclusively on the amount of parts that a certain concept actually has, and the performance of module 1 which detects them. Obviously this number would say nothing about the actual novelty of the creation.
A better approach would be to reuse a metric that has guided the creative process, *semantic similarity*, and reconsider it as a metric for novelty. To illustrate this choice, consider a blend of two concepts that have a high semantic similarity, such as globe and universe. Depending on which image is chosen as a base image, the results would depict a real terrestrial sphere on a pedestal, or a geographical globe in the middle of the universe. In both cases, the result has a low value of novelty, since they are almost identical to the original concepts. In contrast, if two concepts are blended that have a low semantic similarty (such as "universe" and "face"), the chances are higher that the result will depict a concept that does not yet exist, and therefore has a high novelty. Since semantic similarity is measured on a continous scale between 0 and 1, but measures the opposite criteria of novelty, we have to substract its value from 1. The corresponding thresholds are set as follows: Alpha is set to 0.6 and beta is set equal to alpha in order to provide a binary distinction between novel and and not novel results without a marge between them. Results that blend concepts which have a semantic similarity of less than 0.4 (1-0.6) would therefore be considered novel.

Now consider the second metric, *value*:
For the assessment of DUPOND [Mendes and Pereira, 2004], a sentence paraphraser, the metric of value was selectively interpreted as describing meaningfulness and understandability. In

both cases the value was determined by human evaluation. For WASP [Gervas, 2000], a poem generator, the metric value was interpreted as the esthetical qualities of a poem and assessed by a team of volunteers (Pereira et al, 2005). For DIVAGO, the assessment of value was different to the former approaches: DIVAGO is driven by the goal to create a creature that fulfills certain criteria. Within the evolutionary process, a metric called *Relevance* is used to assess intermediate states of the result and guide its development. This metric is reused for the assessment of value. Our system is different to DIVAGO in the sense that we don't possess a metric that guides the development towards a certain goal. Our system has the more general goal to generate images that can be rated as creative. It seems therefore appropriate to follow the approaches taken by the evaluation of WASP and DUPOND, and to assess the metric *value* by a group of human evaluators. We will describe briefly how such a evaluation procedure could be designed.

The group of evaluators consists of 25 volunteers. Each is given the 5 best output images of the system. They are asked to evaluate each of the images on the following three criteria:

- Aesthetic value (is it aesthetically pleasing, surprising, interesting)
- Meaningfulness (Can you estimat a certain meaning in the creation or even different meanings, a richness of meaning? If you can't name the meaning, to what degree does the image have significance, or is it completely pointless to you?)
- Interestingness (How interesting is this image in general, considering both esthetical and semantical aspects)

The three given scores for each image per evaluator are combined to an overal score, which is calculated as the mean of all three scores. The final score for each image is the mean of the given scores of all evaluators for this image.

Since it is not possible to set up such an evaluation study within the given amount of time, we have to ignore the metric of *value* here and base our evaluation on the metric of *novelty*. This has the consequence that the original 14 criteria are pruned down to 6: Criteria 1 and 2 (average novelty and ratio of novel results), criteria 9 and 10 (reinventions), and 11 and 13 (novel results within the set of non-reinventions).
The length of R is not defined, therefore we decided to set it to 5 (the 5 best results that the system outputs in one run).

Finally by grouping the results by criteria, one can propose a summary of conclusions: Our system was able to generate high novelty (from 1 and 2) and no reinventions (9 and 10). It is undoubtable that this evaluation does not represent a reliable assessment of the overall creativity of our system, since one of the two main criteria is omitted. However, it provides a sufficiently detailed evaluation of one essential property, namely *novelty*.

N = 5, Threshold for novelty = 0.6

| Crit. | Meaning | Experiment |
|---|---|---|
| 1 | Average novelty | 0.738 |
| 2 | Novel results/all results | 0.8 (4 of 5) |
| 9 | Reinventions/inspiring set | 0 |
| 10 | Reinventions/all results | 0 |
| 11 | Average novelty of new results | Same as 1 because no reinventions |
| 13 | Novel new results /new results | Same as 2 because no reinventions |

Figure 13: Evaluation on the 6 remaining criteria

# 8 CONCLUSION

As a powerful mechanism for creativity, Conceptual Blending is capable of creating new concepts from known concepts. We showed in concrete steps and modules how the mechanism can be formalized in computational terms and how it can be applied to the visual domain in order to create images that can be rated as creative. We described how the selection process may look like, a topic that has not yet been fully explained in the theory of conceptual blending, and emphasized the fact that knowledge of different kind is crucial to guide the selection process. We think that the architecture of our system is a promising framework for applying conceptual blending to the visual domain. Additional effort in the task related to computer vision will achieve a fully automation of the process. It is also obvious that the quality of the results may be improved when we the size of the inspiring set (the image base) and the amount of knowledge in the knowledge base increases.

# 9 DISCUSSION

When we defined the research objective to generate images that can be rated as creative, we implicitly expected the system to generate images that show some meaningfulness. If the blend was able to create a certain semantic relation between two concepts, which has been before undetected, then the blend has produced a meaningful analogy. The core idea is to associate the unassociable, or stated differently: to find hidden semantic relations in order to surprise and create meaning. However, our system was not built to search explicitly for these types of hidden semantic relations. Rather, they are supposed to emerge when the system searches for formal (visual) analogies. This phenomenon can be seen as an example of the "Eliza Effect", which describes the situation that an artificial system is accredited with understanding while it is actually not the case. Or to put it in terms of W. King (1995): The computer system is perceived as having "intrinsic qualities and abilities which the software controlling the (output) cannot possibly achieve". Instead of relying on pure chance that a hidden relation will emerge, one could also think to create second semantic network of purely hidden relations, called the "creative brain". The "creative brain semantic network" has a significant difference to a "common sense semantic network" such as *Conceptnet*, because it does not rely on common sense but "creative sense". The concepts that are related within the "creative brain" would have generally a long path distance within a common sense network, and are connected via an invented/discovered relation. This relation should not be hardcoded, but inferred from other sources of knowledge. This relation, however, should be understood by the recipient, in order to access the meaning, and therefore its value.

Considering the future work, it is obvious that the most urgent improvement regards the detection of circular shaped subparts. Once this step is automated, the system can be regarded as autonomous. The next step for improvement would then be to gather more knowledge about the subparts in general. Therefore we could take into account the relations that the detected subparts have towards each other: Are they distributed in a horizontal or vertical line (e.g. the lamps on the traffic lights), along a circular line (e.g. the numbers on a clock) or completely randomly (e.g the toppings on a pizza)? The consideration of geometrical distribution of the parts can make the criterion of similar visual properties more subtle and would result in a more advanced way of selecting candidates for blend. Another possibility for improvement is the detection of other geometrical shapes than circles, such as rectangles, triangles etc. Finally we could also go a step further by looking for other image features apart from geometrical shape. SIFT (Scale Invariant Feature Transformation) may be a promising tool for detecting visual similarities between images that are not bound to geometrical shapes but to more subtle local features.

Considering the elaboration of the composed image, there is some room for a better integration of the composed parts. We already enhance the integration by tinting the subparts of the substitute image and by blurring the mask. However, a sophisticated image composition would also consider the lightening conditions within each image. If in the baseimage the light comes from a certain point, resulting in noticeable light reflections and shadows, then it would be consequent to emulate the lightening conditions in the substitute image. This step would change the image in a very subtle way, however, the human perceptory system is very sensible for these kinds of adaptations.

Finally, there is a necessity to provide an objective evaluation of the generated images. As we showed, this is quite a hard task and involves the set-up of a survey with human evaluators. We explained how such a survey might look like. There is another, quite unusual way to achieve the assessment by human evaluators. In the world wide web, several websites exist that provide a platform to exhibit and discuss user made art works. The website *Deviant Art*[2] is an example for this. *Deviant art* has over 26 million members, that submit over 1.4 million "favorites" and 1.5 million comments per day[3], allowing the site to be called "a [free] peer evaluation application" [Mccreight, 2012].

The generated images of our system could be uploaded on this website and be rated by the art-interested public without confessing the fact that they were generated by a computer. This would be a very interesting experiment and a promising contemporary alternative to the usual way of assessing creativity.

[2] `http://www.deviantart.com`

[3] `http://www.youtube.com/watch?v=9O2lRAooUWk`

# References


[Boden, 1991] Boden, M. (1991). *The creative mind. Myths and Mechanisms.* Basic Books, New York.

[Falkenhainer et al., 1989] Falkenhainer, Forbus, and Gentner, D. (1989). The structure mapping engine. Algorithms and examples. *Artificial Intelligence*, 41.

[Fauconnier and Turner, 2001] Fauconnier, G. and Turner, M. (2001). *Conceptual Blending and the mind's hidden complexities.* Basic Books, New York.

[Finke et al., 1992] Finke, R., Ward, T., and Smith, S. (1992). Creative Cognition. Theory, research and applications. *Cambridge, MIT press*.

[Gentner, 1983] Gentner, D. (1983). Structure-Mapping. A theoretical framework for analogy. *Cognitive Science*, 7:155–170.

[Gervas, 2000] Gervas, P. (2000). Wasp: Evaluation of different strategies for the automatic generation of Spanish verse. *AISB-00 Symposium on Artificial Intelligence and Creativity and Cultural Aspects of AI*, pages 33–34.

[Goguen and Harrell, 2004] Goguen, J. and Harrell, F. (2004). Foundations for Active Multimedia Narrative: Semiotic spaces and structural blending. *Interaction Studies: Social Behaviour and Communication in Biological and Articial Systems*.

[Li et al., 2012] Li, B., Zook, A., Davis, N., and Riedl, M. (2012). Goal-Driven Conceptual Blending: A Computational Approach for Creativity. *Proceedings of the Third International Conference on Computational Creativity*.

[McCorduck, 1990] McCorduck, P. (1990). *Aarons Code*. W H Freeman Co.

[Mccreight, 2012] Mccreight, B. M. (2012). A Comparison of Peer Evaluation: The Evaluation App versus DeviantArt. *Purdue University*, pages 33–34.

[Mendes and Pereira, 2004] Mendes, M. and Pereira, C. (2004). Creativity in natural language: Studying lexical relations. *Workshop on Language Resources for Linguistic Creativity*, pages 33–34.

[Miller, 1995] Miller, G. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, 38:39–41.

[Pereira, 2005] Pereira, F. (2005). A Computational Model of Creativity. Universidade de Coimbra, Phd Thesis.

[Pereira and Cardoso, 2002] Pereira, F. and Cardoso, A. (2002). The boat-house visual blending experience. *Proc. to the 2nd Workshop on Creative Systems, ECAI 02*.

[Poincare, 1982] Poincare, H. (1982). *The Foundations of Science: Science and Hypothesis, The Value of Science, Science and Method.* Washington DC.

[Ritchie, 2001] Ritchie, G. (2001). Assessing creativity. *Proceedings of the AISB 01 Symposium on Artificial Intelligence and Creativity in Arts and Science*, pages 3–11.

[Rother et al., 2004] Rother, C., Kolmogorov, V., and Blake, A. (2004). Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Trans. on Graphics (TOG)*, 23:309–314.

[Speer and Alonso, 2007] Speer, H. and Alonso, J. (2007). Conceptnet 3: a flexible, multilingual semantic network for common sense knowledge. *Recent Advances in Natural Language Processing*.