Let's Figure This Out: A Roadmap for Visual Conceptual Blending

João M. Cunha, Pedro Martins and Penousal Machado
CISUC, Department of Informatics Engineering
University of Coimbra
{jmacunha,pjmm,machado}@dei.uc.pt

Abstract

The computational generation of visual representation of concepts is a topic that has deserved attention in Computational Creativity. One technique that is often used is visual blending – using two images to produce a third. However, visual blending on its own does not necessarily have a strong conceptual grounding. In this paper, we propose that visual conceptual blending be used for concept representation – a visual blend complemented by conceptual layer developed through elaboration. We outline a model for visual conceptual blending that can be instantiated in a computational system.

Introduction

It is often said that an image is worth more than a thousand words. Such is aligned with the views from Petridis and Chilton (2019) who state that visual advertisements – specifically when visual metaphors are used – are much more persuasive than plain advertisements or text alone when conveying a message.

Visual metaphors result from a process of visual blending grounded on a conceptual level. The technique of Visual Blending (VB) consists in merging two or more visual representations (e.g. images) to produce new ones. On the other hand, Conceptual Blending consists in integrating two or more mental spaces – knowledge structures – in order to produce a new one, the blend(ed) space (Fauconnier, 1994; Fauconnier and Turner, 2002). When CB and VB are used together, the process can be referred to as Visual Conceptual Blending (Cunha, Martins, and Machado, 2018a), which we consider to have an important role in the production of Visual metaphors.

Visual metaphors are sometimes hard to decode by humans – Petridis and Chilton (2019) report 41.3% accuracy in correctly interpreting visual metaphors. Getting computational systems to understand them is even harder – most of the success comes not from analysing images but from using the accompanying text (Petridis and Chilton, 2019). On the other hand, there is promising work regarding computational approaches to the production of Visual Metaphors.

Cunha and Cardoso (2019) highlight the importance of having a conceptual ground for producing visual blends, creating a connection between conceptual blending and visual blending. This idea was already referred by several authors

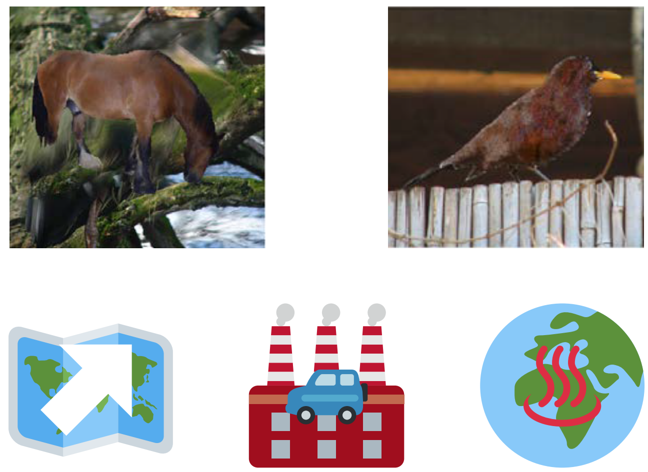

Figure 1: Examples of visual blends produced by Vismantic for *bird/horse* (top row) and Emojinating for *freedom, car factory* and *sun* (bottom row).

but, as far as we know, no concrete computational model has been proposed.

Existing research that relates to the proposal of a visual conceptual blending model somehow comes short of achieving such goal. Karimi et al. (2018) initially present their work as a computational model for generating visual conceptual blends in the domain of sketching. However, the core of the model is more related to conceptual shifts – retrieving sketches similar to an initial one – than with visual blending, which is later presented as a possible application and not intended as an automatic process.

Chilton, Petridis, and Agrawala (2019) propose a workflow for producing Visual Blends, which allows users to generate visual blends collaboratively. The workflow is composed of three main steps: brainstorming, synthesis, and iteration. The user is responsible for finding suitable related concepts (association), retrieving appropriate images that represent the gathered concepts and annotating them in regards to the shape of their elements. The system then finds matches based on shape between the annotated images and combines them into a blend. The user is then responsible for evaluating the results. The process can be repeated until the

Proceedings of the 11th International Conference on Computational Creativity (ICCC'20)
ISBN: 978-989-54160-2-8

Figure 2: Animal visual blends. All blends were created by Arne Olav (gyyporama.com), with the exception of elephaneleon.

users are satisfied. We consider this system as more close to a creativity support tool than a creative system.

Xiao and Linkola (2015) propose a workflow for generating visual compositions to express certain meanings, composed of three tasks: (i) finding photos of the subject and message; (ii) preprocessing photos; and (iii) applying visual operations (juxtaposition, replacement and fusion) to produce blends (see Fig. 1).

Other systems exist in which the conceptual layer can be considered reduced as they often rely on a mere mapping between the input concepts and the visual representations used in the blend (Cunha, Martins, and Machado, 2018a; Zhao et al., 2020). Nonetheless, the system by Cunha, Martins, and Machado (2018a) can, in part, be considered an exception as it provides a mechanism for extending to related concepts.

We propose that a visual conceptual blending should not only result in a visual blend produced for a given concept but instead be complemented by a much more developed conceptual layer (e.g. accompanied by other data such as a name or a description).

Despite providing valuable clues on the direction towards a possible model on visual conceptual blending, these systems cannot be considered as one. In our opinion, they fail to address several topics that we believe are important when building visual conceptual blends.

In this paper, we aim to take a step closer to outlining a model for visual conceptual blending that can be instantiated in a fully operational computational system. Nonetheless, our main goal is to provide a roadmap rather than a final blueprint, providing a broad description that mentions all the topics that we deem important to build such a model. The authors admit that this roadmap is most likely incomplete and will need to be improved in future iterations.

From Metaphors to Visual Blending

According to Peterson (2018), in a metaphor a source domain is recalled for comparison, and aspects of its identity are mapped onto a target domain.

When it comes to visual metaphors, they consist in the combination of objects that establish a similar comparison. However, their interpretation is harder as they are not direct in conveying messages (Petridis and Chilton, 2019), often lacking visual cues on what is the source and the target (Forceville, 1994). Upon observing a visual metaphor, one is able to recognise an object but at the same time notices something strange about it, causing a search for meaning (Chilton, Petridis, and Agrawala, 2019).

Despite not all visual blends being visual metaphors, most research is conducted in relation to advertisement and focuses on them (e.g. Phillips and McQuarrie, 2004; Gkiouzepas and Hogg, 2011).

Components and Types of Visual Blending

Chilton, Petridis, and Agrawala (2019) define a visual blend as having the following properties:

- Two concepts are given as input and each is mapped to an object that visual represents or symbolises it;
- The visual blend is an object that integrates the two initial objects, in a way that they are still recognisable and allow the user to infer an association between the concepts.

Regarding blend types, one categorisation was done by Phillips and McQuarrie (2004), who propose three types of blending of increasing complexity: juxtaposition (depict both source and target), fusion (the domains are combined) and replacement (one of the domains is omitted).

Peterson (2018) presented an expansion to this typology: identification – one domain pictorial, other textual; pairwise juxtaposition – both entities complete and separate, equal to juxtaposition by Phillips and McQuarrie (2004); categorical juxtaposition – source amidst target set, relates to that category concept; replacing juxtaposition – one entity breaks a set of selfsame entities, replacing one instance; replacement – one entity is absent and must be imagined by the viewer using contextual cues, equal to replacement by Phillips and

McQuarrie (2004); replacing fusion – part of one entity is replaced by another entity or part of it; and fusion – two entities are fused together to form a hybrid.

This typology seems easier to employ and to better match what is done by several authors working on visual blending. For example, Cunha, Martins, and Machado (2018a) use the term "replacement" but what their system performs – using an emoji to replace part of another – is more aligned with "replacing fusion" as defined by Peterson (2018).

Analysis to Visual Blends

According to Pollak et al. (2015), there are still many open questions regarding the production of blends. By investigating human creations and identifying patterns, it is possible to address these questions and possibly find a direction for the blending process, eventually allowing the automated generation of blends (Pollak et al., 2015). Joy, Sherry Jr, and Deschenes (2009) conducted an analysis of blends based on human perception by analysing conceptual blending in advertising. Bolognesi, van den Heerik, and van den Berg (2018) built a corpus of visual metaphors that have been analysed and annotated on different dimensions of meaning. Petridis and Chilton (2019) focus on how people interpret visual metaphors of advertisements and identified causes for misinterpretation.

For our work, the most interesting example of blend analysis was conducted by Martins et al. (2015), who conducted an online-survey questionnaire in which participants were asked to evaluate criteria assumed to be related to the quality of blends. Martins et al. (2015) used visual blends between two animals (see Fig. 2) and tried to identify what humans perceive as a good blend. These blends used fusion and were focused on perceptual features, e.g. color, texture, or pattern.

Upon analysing the blends (see Fig. 2), one observes that colour cannot be considered the main reason for conducting the blend – i.e. animals are not blended on the basis of similar colour – but as a way to produce a good blend by achieving a fully integrated blend. Nonetheless, in some blends colour alignment of the input animals seems occur (e.g. *pengwhale* or *guinea lion*). In the same way proportion is also not the ground for blending, as several examples exist of strange proportion between head and body (e.g. *snorse*). It leads to the conclusion that the selection of the input animals was conducted without any apparent reason or conceptual grounding. Regarding the mapping that leads to the blend, one can see that it is mostly based in element category similarity (e.g. head of the snake is mapped to the head of the horse). Nonetheless, in (Martins et al., 2015) special attention is given to elaboration: name building and context creation.

Another example of blend analysis is described by Chilton, Petridis, and Agrawala (2019), who stated that they observed blend examples and tested theories to come up with a design patter – they identified shape as a particularly important feature in visual blending. Based on this, they developed a workflow for producing visual blendings based on an abstract structure: blend two objects that have the same basic shape but other identifying visual features. This example contrasts with the one from Martins et al. (2015) as they use a completely different feature. In addition, whereas Martins et al. (2015) only used blends of animals (fusion blend type), Chilton, Petridis, and Agrawala (2019) analysed visual blends of objects based on replacing fusion type. Similar studies are needed with other types of blend.

Roadmap for Visual Conceptual Blending

In this paper we present a model for the production of visual blendings with a strong conceptual grounding. In a process of visual conceptual blending, despite the output being a visual blend, it does not merely consist in the task of producing a merge of two initial visual representations. Instead, core of the process has to do with conceptual reasoning, which serves as base for the actual process of visual blending. This contrasts with the description given by Chilton, Petridis, and Agrawala (2019) for the constituents of a visual blend, presented in the previous section.

In visual conceptual blending, the focus is not the transformational task of mixing two images but the whole process of producing visual blends that are based on a conceptual reasoning and present themselves as a result of a knowledge-based process. In fact, from our perspective, the output of a process of visual conceptual blending is not only an image but also a set of conceptual elaborations. A visual conceptual blending has context, it is grounded on a justification which should indicate the relevance of the blend. It can also be given a name that may not even be aligned with the original concept.

In this section, we outline a model for the production of visual conceptual blends. Our roadmap is composed of four main stages: (i) Conceptualisation; (ii) Visual Blending; (iii) Quality Assessment; and (iv) Elaboration.

Despite presenting it as a series of stages that may seem to occur in a linear sequence, the reader should understand that order may vary, not being fixed and allowing the repetition of some of the stages. These stages are somehow aligned with the three operations proposed by Fauconnier and Turner (2002): composition, completion, and elaboration.

Conceptualisation

By just focusing on the visual blending process, we can generate an infinite number of possibilities. However, these are not guaranteed to be grounded on knowledge. For example, for the visual blend between a snake and a horse, instead of the logical category-category mapping between heads (seen in Fig. 2) it would be possible to produce a blend in which head of the snake replaces the tail of the horse. However, such blend would have a very low conceptual grounding as no apparent logical mapping was performed. Conceptualisation is what distinguishes mere generation from something with a strong conceptual grounding (resultant from a process of reflection) and consequently visual blending from visual conceptual blending.

Conceptualisation can occur in at least two stages: selection of input concepts and mapping between previously given ones. Most examples described in this paper fall under the latter case – the input concepts to use in the blend are previously chosen. However, in a general model an initial step may be to identify potential candidates for a visual conceptual blending (Gonçalves et al., 2015). As such the topics addressed in this section can mostly be applied to these two situations.

In fact, the process of conceptualisation may lead to the retrieval of related concepts for the production of the visual blend (e.g. Veale, 2012; Cunha, Martins, and Machado, 2018a). In such case, there may not be a direct relation between the origin concept and the visual blend. This is specially evident when the origin concept consists in one word – it is necessary some sort of expansion to provide a foundation for a processing of visual blending to occur. In these cases, the origin concept can be visually represented by resorting to related concepts, e.g. *freedom* represented using a related concept "universal right" (see Fig. 1). In turn, the interpretation of the resultant visual blend can even lead to a third concept, for example "travel the world". Therefore, the levels of conceptualisation of a visual blend can vary. In fact, the process of conceptualisation can reach high degrees of complexity – e.g. using a process of conceptual blending based on structural alignment techniques to produce analogies from structures such as mental spaces (see Fig. 4) – such was used by Cunha et al. (2017).

On the other hand, a process of visual conceptual blending can have different motivations and therefore different goals. For example, the process can be used for concept representation, in which case a literal representation may be preferred. Another possibility may be the production of visual metaphors, in which case the goal will be more creative.

In the end, the conceptualisation stage consists in answering the question: what is behind the blend? How this question is approached depends on the starting point. For example, if we already have a two word concept it is more related to how we blend the two concepts – finding a justification for a blend. If the starting point is a one word concept, we face a somehow open search for potential blends – which is good if we have enough knowledge.

Several characteristics can motivate the process of blend, e.g. conceptual features (e.g. name or affordances) or perceptual features (e.g. shape or colour).

Grounding the blend: perceptual features One way of grounding the blend is by using perceptual features, e.g. shape or colour. The usefulness in perceptual features is especially relevant when these include prototypical elements (Johnson, 1985) – i.e. what most identifies a given concept (e.g. the nose and the tails in a pig). An example is the work by Karimi et al. (2018) in which blend possibilities are found using a process of conceptual shift based on shape comparison.

Obviously, these characteristics are very dependable on the representation used – e.g. if only black and white images are to be used, colour loses relevance. The mappings based on perceptual features always depend on the situation.

Grounding the blend: affordances Another way of finding blend possibilities is related to affordances and their modelling using, for example, image schemas. Such may help in guiding how the visual blending should be conducted – e.g. using the schema CONTAINMENT with icons of "money" and "building" to represent *bank* (Cunha, Martins, and Machado, 2018b; Falomir and Plaza, 2019).

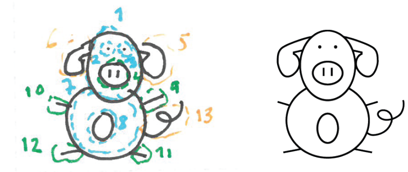

Figure 3: On the left is a the representation drawn with the elements identified; On the right is the result of the conversion into fully scalable vector graphic (Cunha et al., 2017)

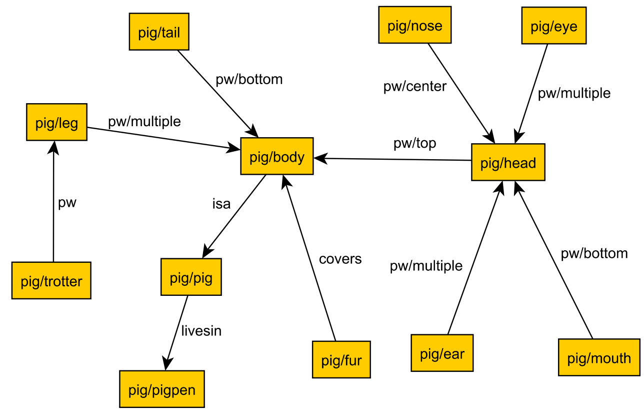

Figure 4: Mental space for pig based on the visual representation

Grounding the blend: naming A third possibility has to do with the name – e.g. finding homophones such as "waste of money" and "waist of money". As an example, Veale and Al-Najjar (2016) explore the invention of colour names.

Visual Blending

Existing visual blending systems can be divided into two groups based on the type of rendering (see Fig. 1): photorealistic, e.g. (Xiao and Linkola, 2015), and non-photorealistic, e.g. (Cunha et al., 2017). These two types have great differences in terms of how the visual blending process occurs. A photorealistic visual blending may require computer vision and image processing techniques, whereas a non-photorealistic visual blending that uses fully scalable vector graphic is much easier to conduct (Cunha et al., 2017).

In either case, a process of visual blending involves two main decisions: which objects to combine and how to combine them.

Connection between Conceptual and Visual Most of the visual blending examples that are grounded on a process of conceptual blending consist in a simple visualisation of the blend, e.g. (Pereira and Cardoso, 2002). An exception can be seen in (Cunha et al., 2017), in which two types of net-

Proceedings of the 11th International
Conference on Computational Creativity (ICCC'20)
ISBN: 978-989-54160-2-8

work structures were used: one corresponding to the mental spaces of the input concepts and another corresponding to the visual structure of the visual representation (see Figs.3 and 4). The two types of structure were aligned to produce visual conceptual blends. In this case, then system was considered a hybrid blender, as the blending process starts at the conceptual level and ends at the visual one. However, this situation is uncommon, as in most cases it is not possible to align the conceptual layer with the visual one – such data would have to be manually built. One possibility would rely on an analysis of the images to produce a network structure (structure extraction). This is more or less easy to implement in fully scalable vector graphics but on raster images it would have to use techniques such as concept detection (Zhou, Jagadeesh, and Piramuthu, 2015).

In any case, in the same way the visual level is based on what is produced on the conceptual level, the conceptual level also needs to take into account the character and features of the representations being used.

Questions of semiotics In addition to the simple exchange of parts, there are several aspects that have to be taken into consideration. Cunha et al. (2015) address this issue and provide some guidance on how color, shape and other visual aspects may affect meaning. Moreover, one should also consider elements such as modifiers (Cohn, 2007) and bear in mind that there are cultural differences that ultimately have an impact how a visual blend is interpreted (Cunha et al., 2015).

Type of Blend Each type of blend is suitable to different types of concepts and visual representations. As such, the choice of which blend type to use should take certain aspects into consideration. First, it should consider the relationship between the categories of the concepts being blended. For example, it is completely different blending “dinosaur” with “park” and “dinosaur” “fish”. The former case involves an animal and a location, which makes it more suitable to have a juxtaposition. In the latter case, both concepts are animals and, as such, a fusion might be more appropriate.

Then, since the process of blending involves visual representations (e.g. icons), the appropriateness of blend type also varies depending on type of representation being used. For example, in the “dinosaur” “fish” the animals are very different and that will have an impact on how the blend is conducted. Moreover, it is completely different to blend two representations that show the full body of the animal and one that shows a full body and another that only shows the head. For the blends shown in Fig. 2, the author mostly likely had to carefully select the images that better matched one another.

Quality Assessment

While producing blend, it is important to have a measure of quality. If we have a system that produces several individuals, a measure of quality is crucial to identify good solutions. In certain situations the blend production can be considered an open-ended problem, in which case including the user in the cycle may provide some advantages. Nonetheless, several types of quality assessment exist – some may be more suitable for certain goals than others.

In fact, Martins et al. (2015) poses several questions regarding quality assessment: “How ‘semantically far’ should the input spaces be to produce a good blend?”, “Is there a correlation between the quality of blends and the number of elements for projection?” or even “Are all the optimality principles required to produce good blends?”. In this section, we present some types of quality measures that can be used to assess how good a blend might be.

Argumentation Confalonieri et al. (2015) proposed the use of argumentation to evaluate and iteratively refine the quality of blended computer icons. The authors introduced a semiotic system, which was based on the idea that signs can be combined to convey multiple intended meanings. Despite this, no evidence of a possible implementation was provided.

Optimality Principles Fauconnier and Turner (1998) proposed a list of optimality principles that can guide the process of conceptual blending. These principles are not trivial to computationally model and are normally used at the conceptual level. Nonetheless, it is also possible to use them to validate the blend on the visual level, as Kowalewski (2008) demonstrated by analysing the formation of logos and product names in terms of usage of optimality principles. Even though these principles are considered as responsible for generating consistent blends (Martins et al., 2015), they should not be regarded as “rigid laws” but as flexible guidelines (Kowalewski, 2008). We provide a description of these principles below:

- Integration: the blend must constitute a tightly integrated scene that can be manipulated as a unit. It should be a coherent, self-contained and unified structure, and not a loosely knit combination of random element (recognized as a whole). Integration is identified by Martins et al. (2015) as the most important principle;
- Topology: the elements projected into the blend should maintain the same neighbourhood relations as in the input space. Even though Martins et al. (2015) indicate that topology as not relevant, according to Kowalewski (2008) it can be useful for example in terms of spacial organisation by placing elements in the blend according to the configuration of one of the input visual representations (e.g. maintaining the existence of a central element, laying new elements according to center-periphery scheme).
- Web: the blend as a unit must maintain the web of appropriate connections to the input spaces, so that an event in one of the input spaces implies a corresponding event in the blend.
- Unpacking: this principle takes the perspective of the reader and consists in the easiness of reconstructing the inputs and the network of connections from the blend. The input concepts should be recognisable from the elements of the blend as the input visual representations or parts of them. One example of this can be seen in the use of prototypical parts in the blend.
- Relevance (or Good Reason): if an element appears in the blend it should have some kind of significance/meaning

(e.g. links to other spaces). This is easy to observe in the application of a color to a blend should be based on the input visual representations – e.g. using green from a snake in *snorse*.

Two other principles are “Intensifying Vital Relations” and “Maximising Vital Relations”. However, in this context we could not provide a clear usefulness for them. In addition to being sometimes vague and difficult to implement, not all the principles are compatible with each other (Martins et al., 2015). Moreover, choosing some over others may lead to more or less creative blends (Martins et al., 2016).

Visual Analysis Assessing quality can also concern visual aspects. Two examples are: overall complexity and area exchanged (Cunha et al., 2020). It is important to mention that some aspects are easier to apply in a visual blending with layered images. For raster images other aspects may be more appropriate

User Perception Despite the importance of all the topics already mentioned, the quality of visual blend will always depend on user perception and interpretation, and may perform poorly even if the blend is conceptually grounded. Providing a way for the user to interact with the system would make it so that improvements could be made to increase the likelihood of a correct interpretation. A method used by some systems (e.g. Cunha et al., 2019) is Interactive Evolutionary Computation, which consists in including the user in the task of fitness assignment and evolving solutions that match their preference.

Elaboration

A big part of the conceptual process may occur after the visual blending is done – consisting of an elaboration. This elaboration and consequent interpretation may in turn serve to provide justification for the previously done visual blend and also as a way to improve it – resulting in a return to a previous stage for a new iteration.

Naming One example of elaboration is the production of names. Pollak et al. (2015) presented a prototype for name generation based on an investigation focused on the principles of creating lexical blends based on visual blends (blended animals). Pollak et al. (2015) identified the following mechanisms used in name formation: L1-concatenation blends; L2-portmanteaux (e.g. rabbear for rabbit and bear); L3-blending based on visible characteristics; L4-blending using background knowledge and L5-bisociative blends (e.g. mickey the bear for mouse and bear). These techniques can be used for other blends that do not use animals.

Descriptions In addition to names, there is also the potential to produce descriptions based on the visual blend. Techniques such as image captioning (Feng et al., 2019) may be used for this purpose. Ideally, a system that produces descriptions could produce an elaboration on the context of the blend. For example, mixing two animals leads to questioning the context of the hybrid animal: Where does it live? What does it eat? How does it behave in relation to other animals? All theses questions would need to be addressed

	Steinbrück (2013)	Xiao and Linkola (2015)	Ha and Eck (2017)	Cunha et al. (2017)	Karimi et al. (2018)	Cunha et al. (2018a)	Chilton et al. (2019)	Zhao et al. (2020)
creativity-support tool					x		x	x
creative system	x	x	x	x		x		
type of blend	R	JRF	F	F	J	JRF	R	JRF
guide of blend	Sha	Con	Con	Con	Con	Con	Sha	Con

Figure 5: Analysis to existing visual blending approaches (R = replacement, J = juxtaposition, F = fusion, Sha = shape, Con = conceptual)

using a process of conceptual blending by getting characteristics from the two mental spaces. An example can be observed in the concept *clown fish*: does it live in the sea and looks like a clown? does it live in a circus and looks like a fish? Obviously, one of the situations has a higher likelihood, which makes it more plausible; but the surprising nature of the other option makes it so that in terms of creativity it has much more potential.

Moreover, a creative system like the one we are proposing would have great advantages in providing the user with explanations for the produced blends. The descriptions can be seen as such and used to make the process of blending clearer to the user (Cook et al., 2019).

Towards Implementation

In this section, we briefly analyse existing approaches using the topics presented in the previous section. Then, we provide some guidelines that we believe should be taken into account when building a general model for visual conceptual blending.

Analysis to existing blend systems

In order to summarise existing approaches, we conducted an analysis to several systems that address visual blending (see Fig. 5). The analysis was made in terms of type of system (creativity-support tools vs creative system), type of blend (replacement, juxtaposition or fusion) and guide of blend (shape or conceptual). The analysed approaches were: Steinbrück (2013); Xiao and Linkola (2015); Ha and Eck (2017); Cunha et al. (2017); Karimi et al. (2018); Cunha, Martins, and Machado (2018a); Chilton, Petridis, and Agrawala (2019); Zhao et al. (2020). None of them addresses all the topics we mention on the model.

General Model

Having analysed existing systems, we now present a set of aspects that, in our opinion, will be key in implementing a general model for visual conceptual blending.

Modularity Most of the systems described before work in an individual way with no connection to others. An exception is Vismantic (Xiao and Linkola, 2015), which is integrated in a platform for workflow management – ConCreTeFlows. Martins et al. (2019) focus on this platform and present an example of how it can be used to develop CC software components that can shared, used and reused to produce complex computational pipelines. We believe that an implementation of a general model for visual conceptual blend will profit from using such modular approach, allowing multiple users to contribute to the system.

Multi-approach In addition to having several modules that deal with different tasks, as we have seen earlier, there are several methods that can be employed for each of the tasks (e.g. conceptualisation can be based in perceptual features, affordances, etc.). The suitability of these methods often depends on the type of problem at hands (i.e. the characteristics of the blend) and, as such, no optimal approach exists. A solution to this multi-approach situation is to follow a similar strategy to the one presented by Cardoso et al. (2015) – using a global workspace and a number of components that compete for access to it. Each component could be seen as an agent. At each time, the agent that is able to produce the most relevant output is given access to the workspace. This would consist in having solutions being produced by each of the agents and finding the best.

User centred The quality of a visual blending always depends on user perception, thus being of open-ended nature. As such, the user should be viewed as having a central role. The modular approach suggested earlier is obviously dependent on having a user interacting with the platform to build the a pipeline of components. We go one step further and propose that the user should also have an active role in producing the visual blends.

First, the interaction with the user has great potential to be explored as it can be used to iteratively improve the quality of the blend, both visually and conceptually. This would consequently have an effect on which approach is used at each task, depending on the user evaluation. Moreover, the user would guide the blend production in terms of improving second-order features (e.g. color) or even extending the conceptual reach when no blends can be produced with the existing knowledge.

Another possibility is to provide the user with a way of selecting the creativity degree they want for the blend – low creativity resulting in literal representations and high creativity in more metaphorical results.

Conclusion and Future Work

In this paper, we focused on visual conceptual blending. We started by providing the reader with a general view on current research on visual blending. Then we presented a proposal of a model for the production of visual conceptual blends. This model can be instantiated into a modular system, in which the different stages of blend production occur in an iterative manner, allowing the user to go back to improve the blend and its elaboration. Future developments concern the implementation of the proposed model, as well as the establishment of collaborations with researchers who develop work in areas related to the identified topics.

Acknowledgements This work is partially funded by the FCT - Foundation for Science and Technology, I.P., within the scope of the project CISUC - UID/CEC/00326/2020 and by European Social Fund, through the Regional Operational Program Centro 2020, and under the grant SFRH/BD/120905/2016.

References

Bolognesi, M.; van den Heerik, R.; and van den Berg, E. 2018. Vismet 1.0: An online corpus of visual metaphors. In Steen, G. J., ed., *Visual Metaphor*. John Benjamins Publishing Company. chapter 4.

Cardoso, A.; Martins, P.; Assunção, F.; Correia, J.; and Machado, P. 2015. A distributed approach to computational creativity. In *Intelligent Distributed Computing IX - Proceedings of the 9th International Symposium on Intelligent Distributed Computing - IDC'2015, Guimarães, Portugal, October 2015*, 3–12.

Chilton, L. B.; Petridis, S.; and Agrawala, M. 2019. Visiblends: A flexible workflow for visual blends. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*, 172.

Cohn, N. 2007. A visual lexicon. *Public Journal of Semiotics* 1(1):35–56.

Confalonieri, R.; Corneli, J.; Pease, A.; Plaza, E.; and Schorlemmer, M. 2015. Using argumentation to evaluate concept blends in combinatorial creativity. In *Proceedings of the Sixth International Conference on Computational Creativity*, 174–181.

Cook, M.; Colton, S.; Pease, A.; and Llano, M. T. 2019. Framing in computational creativity-a survey and taxonomy. In *ICCC*, 156–163.

Cunha, J. M., and Cardoso, A. 2019. 3.2 from conceptual blending to visual blending and back. In Besold, T. R.; Gervás, P.; Gius, E.; and Schulz, S., eds., *Computational Creativity Meets Digital Literary Studies (Dagstuhl Seminar 19172) – Dagstuhl Reports*, volume 9. 92.

Cunha, J. M.; Martins, P.; Cardoso, A.; and Machado, P. 2015. Generation of concept-representative symbols. In *Workshop Proceedings of the 23rd International Conference on Case-Based Reasoning (ICCBR-WS 2015)*. CEUR.

Cunha, J. M.; Gonçalves, J.; Martins, P.; Machado, P.; and Cardoso, A. 2017. A pig, an angel and a cactus walk into a blender: A descriptive approach to visual blending. In *Proc. of the Eighth Int. Conf. on Computational Creativity*.

Cunha, J. M.; Lourenço, N.; Correia, J.; Martins, P.; and Machado, P. 2019. Emojinating: Evolving emoji blends. In *Proc. of the 8th Int. Conf. on Computational Intelligence in Music, Sound, Art and Design*. Springer.

Cunha, J. M.; Lourenço, N.; Martins, P.; and Machado, P. 2020. Emojinating Co-Creativity: Integrating Self-Judgement and Context-Adaptation. In *Proceedings of the Eleventh International Conference on Computational Creativity*.

Cunha, J. M.; Martins, P.; and Machado, P. 2018a. How shell and horn make a unicorn: Experimenting with visual blending in emoji. In *Proc. of the Ninth Int. Conf. on Computational Creativity, Salamanca, Spain, June 25-29, 2018.*

Cunha, J. M.; Martins, P.; and Machado, P. 2018b. Using image schemas in the visual representation of concepts. In *Joint Proceedings of the Workshops C3GI: The 7th International Workshop on Computational Creativity, Concept Invention, and General Intelligence ISD4: The 4th Image Schema Day, and SCORE: From Image Schemas to Cognitive Robotics, Bozen-Bolzano, Italy, December 13-15, 2018.* CEUR.

Falomir, Z., and Plaza, E. 2019. Towards a model of creative understanding: deconstructing and recreating conceptual blends using image schemas and qualitative spatial descriptors. *Annals of Mathematics and Artificial Intelligence*.

Fauconnier, G., and Turner, M. 1998. Conceptual integration networks. *Cognitive Science* 22(2):133–187.

Fauconnier, G., and Turner, M. 2002. *The Way We Think*. New York: Basic Books.

Fauconnier, G. 1994. *Mental Spaces: Aspects of Meaning Construction in Natural Language*. New York: Cambridge University Press.

Feng, Y.; Ma, L.; Liu, W.; and Luo, J. 2019. Unsupervised image captioning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4125–4134.

Forceville, C. 1994. Pictorial metaphor in advertisements. *Metaphor and symbol* 9(1):1–29.

Gkiouzepas, L., and Hogg, M. K. 2011. Articulating a new framework for visual metaphors in advertising. *Journal of Advertising* 40(1):103–120.

Gonçalves, J.; Martins, P.; Cruz, A.; and Cardoso, A. 2015. Seeking divisions of domains on semantic networks by evolutionary bridging. In *ICCBR (Workshops)*, 113–122.

Ha, D., and Eck, D. 2017. A neural representation of sketch drawings. *arXiv preprint arXiv:1704.03477*.

Johnson, R. 1985. Prototype theory, cognitive linguistics and pedagogical grammar. *Working Papers in Linguistics and Language Training* 8:12–24.

Joy, A.; Sherry Jr, J. F.; and Deschenes, J. 2009. Conceptual blending in advertising. *Journal of business research* 62(1):39–49.

Karimi, P.; Maher, M. L.; Grace, K.; and Davis, N. 2018. A computational model for visual conceptual blends. *IBM Journal of Research and Development*.

Kowalewski, H. 2008. Conceptual blending and sign formation. *The Public Journal of Semiotics* 2(2):30–51.

Martins, P.; Urbancic, T.; Pollak, S.; Lavrac, N.; and Cardoso, A. 2015. The good, the bad, and the aha! blends. In *6th International Conference on Computational Creativity, ICCC 2015*.

Martins, P.; Pollak, S.; Urbancic, T.; and Cardoso, A. 2016. Optimality principles in computational approaches to conceptual blending: Do we need them (at) all? In *Proceedings of the Seventh International Conference on Computational Creativity*.

Martins, P.; Oliveira, H. G.; Gonçalves, J. C.; Cruz, A.; Cardoso, F. A.; Žnidaršič, M.; Lavrač, N.; Linkola, S.; Toivonen, H.; Hervás, R.; et al. 2019. Computational creativity infrastructure for online software composition: A conceptual blending use case. *IBM Journal of Research and Development* 63(1):9–1.

Pereira, F. C., and Cardoso, A. 2002. The boat-house visual blending experience. In *Proceedings of the Symposium for Creativity in Arts and Science of AISB 2002*.

Peterson, M. O. 2018. Aspects of visual metaphor: an operational typology of visual rhetoric for research in advertising. *International Journal of Advertising* 38(1):67–96.

Petridis, S., and Chilton, L. B. 2019. Human errors in interpreting visual metaphor. In *Proceedings of the 2019 ACM SIGCHI Conference on Creativity and Cognition, C&C 2019, San Diego, CA, USA, June 23-26, 2019.*, 187–197.

Phillips, B. J., and McQuarrie, E. F. 2004. Beyond visual metaphor: A new typology of visual rhetoric in advertising. *Marketing theory* 4(1-2):113–136.

Pollak, S.; Martins, P.; Cardoso, A.; and Urbancic, T. 2015. Automated blend naming based on human creativity examples. In *23rd International Conference on Case-Based Reasoning (ICCBR 2015) – Workshop on Experience and Creativity*.

Steinbrück, A. 2013. *Conceptual blending for the visual domain*. Ph.D. Dissertation, Master's thesis, University of Amsterdam.

Veale, T., and Al-Najjar, K. 2016. Grounded for life: creative symbol-grounding for lexical invention. *Connection Science* 28(2):139–154.

Veale, T. 2012. From conceptual "mash-ups" to "badass" blends: A robust computational model of conceptual blending. In Maher, M. L.; Hammond, K.; Pease, A.; Rafael Pérez; Ventura, D.; and Wiggins, G., eds., *Proceedings of the Third International Conference on Computational Creativity*, 1–8.

Xiao, P., and Linkola, S. 2015. Vismantic: Meaning-making with images. In *Proc. of the 6th Int. Conf. on Computational Creativity, ICCC-15*.

Zhao, N.; Kim, N. W.; Herman, L. M.; Pfister, H.; Lau, R. W.; Echevarria, J.; and Bylinskii, Z. 2020. Iconate: Iconate: Automatic compound icon generation and ideation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*.

Zhou, B.; Jagadeesh, V.; and Piramuthu, R. 2015. Conceptlearner: Discovering visual concepts from weakly labeled image collections. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1492–1500.