Vision, Language, Art Workshop

You are warmly invited to attend a workshop on Vision, Language and Art. To be held in Bath, UK, 12-13 September 2024.

Significant progress is being made in visual computing by linking vision models with large language models. This vision-language relationship can be used for tasks such as "image to text" (generate text given an image) and "text to image" (generate an image from text), and we are witnessing increasingly powerful applications of these tasks.

For example, generative models such as DALL-E or SORA can produce impressive results, while foundation models such as Flamingo are capable of impressive text completion given text prompts. More generally, language in vision is proving capable of addressing otherwise difficult problems such as zero-shot learning. However, models are expensive to train and, despite the successes, it is not at all clear that they generalise well to visual artwork.

The inclusion of art and humanities pushes at the fundamental assumptions of even the most powerful vision models. Art provides a wealth of examples that current models fail to adequately generalise to. Recognition, for example, tends to be limited to photographs, or to art that looks somewhat photographic. In contrast, all current machines have problems recognising objects in abstract art, the art of children, art of the ancient world, and many other types of art. Some art is deliberately ambiguous - the rabbit/duck illusion is just one example. Other times the same object may be depicted very differently in different images -- Mickey Mouse has changed appearance (and clothes) over time, some depictions of him are very abstract.

Within this context, a few challenging problems and applications arise such as:
* Recognition of out-of-distribution objects, given very limited training data,
* Automatically reading manuscripts, Assyrian reliefs, hieroglyphics, ...
* Create art using spoken language,
* Indexing heterogeneous databases (paintings, sculptures, masks, stone tools, texts, ...).

Importantly, the field is not without ethical issues. In fact, artists are concerned that their jobs can be replaced by AI., s. Artists can also recognise their style in AI-generated art, even if they did not agree to their images being used to train AI models. .

Our workshop on Vision, Art and Language aims to be a forum to discuss these topics (and many more as well). We will bring together people with different backgrounds: researchers in computer vision and natural language processing, art historians and artists. With a diverse set of expertise, we believe our workshop will be an exciting opportunity to engage in an insightful conversation on what are the new frontiers of computer vision to understand art, how vision models can be improved with art, and how can artists and historians benefit from AI.

It is an invitation only event, and we seek a good balance between different interest groups. Please respond as soon as possible to help us in that task. We may be able to help you with costs, please contact Peter (maspmh@bath.ac.uk) directly if you wish to find out more about that.

Shen et al. "How much can clip benefit vision-and-language tasks?." arXiv preprint arXiv:2107.06383 (2021).
Alayrac et al. "Flamingo: a visual language model for few-shot learning." Advances in neural information processing systems 35 (2022): 23716-23736.

Radford, et al. 2021. Learning transferable visual models from natural language supervision. arXiv preprint arXiv:2103.00020.