Vector symbolic visual analogies

Navneedh Maudgalya **Bruno Olshausen** **Spencer Kent**

Redwood Center for Theoretical Neuroscience, UC Berkeley

Reasoning via analogy depends on one of the most striking traits of human cognition, namely a facility for conceptual abstraction and the flexible recombination of atomic components in order to describe situations never before encountered. This submission is about our work to capture elements of conceptual abstraction and analogy in visual scenes, using a framework called Vector Symbolic Architectures (VSAs). Vector Symbolic Architectures are a family of theories for cognitive representation in the brain that provide operations for building *datastructures* using high-dimensional vectors (Gayler, 2004; Plate, 2003; Kanerva, 2009) and, in combination with neural networks, suggest a path toward systems that use symbolic computation but also learn from data.

Prior work in the Deep Learning literature has constructed so-called 'visual analogy' problems and leveraged large neural networks with relatively little imposed structure to map an inferred relation between two images onto a third image (Reed et al., 2015). These analogies follow the form `A : B :: C : ?`, where `A` and `B` are two images between which a relation is computed and `C` is an image onto which this relation is applied in order to produce an output image. The method of computing such a relation in these models is extremely simple, namely the subtraction between embedded vector representations for the images `A` and `B`.

We believe this lacks a full appreciation for the nature of analogy, namely the substitution of atomic components which are embedded in larger datastructures. This computation is not well-captured by simple vector subtraction. However, in the framework of VSAs, substitution of atomic components within/between complex datastructures is elegant and highly efficient when the datastructures are constructed by a *multiplicative* operation called *binding*. The operation of unbinding somewhat naturally reveals analogical relations that can be used in various ways, and has been discussed in the VSA literature (Plate, 2000; Kanerva, 2010). Our contribution has been to examine the application of VSAs to modeling structure in visual scenes. Some of our prior work (Kent and Olshausen, 2017) has shown how one can learn a VSA-structured representation of scenes in order to support structured transformations, as we depict in Figure 1.

Using some of these ideas, we have endeavored to compute visual analogies via the binding and unbinding operation of VSAs. This requires learning a mapping from the space of images into the space of symbolic vectors that imposes certain mathematical structure on the vectors. Such a task can be challenging and is a topic of active research in the VSA community. We have taken a visual analogy problem from prior deep learning work (Reed et al., 2015) and made it significantly more challenging, showing that the black-box deep learning approach struggles, ostensibly due to the fact that it does not build a representation with sufficient structure to successfully generalize these analogies to unseen data. The new analogy task is depicted in Figure 2. Our ongoing efforts have been to solve this visual analogy problem using VSAs, and a much more principled conception

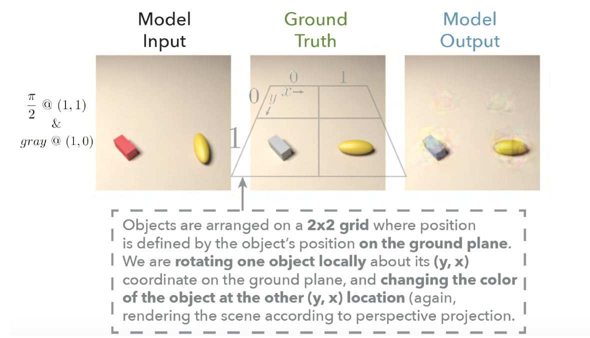

Figure 1: Factorization of position, surface color, and rotation supports structured transformation of visual scenes using VSA operations on latent high-dimensional symbolic vectors

of what constitutes these analogies. This is fairly new work and has not yet born fruit in the sense of solving the analogies depicted in Figure 2 but we are fairly confident that the VSA approach will afford certain advantages over prior deep learning works, based on our prior experience with encoding structured representations of visual scenes. Vector symbolic algebras show how symbolic computation can be reconciled with connectionism, and may pave a path toward robust visual analogies.

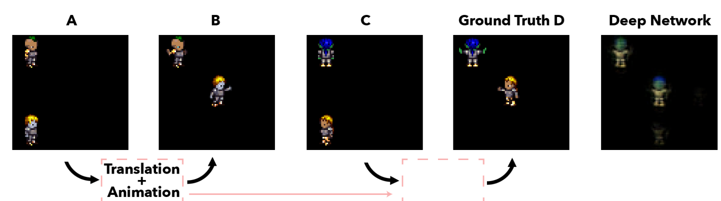

Figure 2: Analogy task using simple animated characters: Images `A` and `B` are related by a translation and animation of each character *separately*. This relation is applied to image `C` to produce image `D`. Deep networks struggle with this type of analogy.

References

Gayler, R. W. (2004). Vector symbolic architectures answer jackendoff's challenges for cognitive neuroscience. *arXiv preprint arXiv:cs/0412059*.

Kanerva, P. (2009). Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. *Cognitive Computation*, 1(2):139–159.

Kanerva, P. (2010). What we mean when we say "what's the dollar of mexico?": Prototypes and mapping in concept space. In *2010 AAAI fall symposium series*.

Kent, S. J. and Olshausen, B. A. (2017). A vector symbolic approach to scene transformation. In *Cognitive computational neuroscience (CCN '17)*.

Plate, T. A. (2000). Analogy retrieval and processing with distributed vector representations. *Expert systems*, 17(1):29–40.

Plate, T. A. (2003). *Holographic reduced representation: Distributed representation of cognitive structure*. CSLI Publications, Stanford, CA.

Reed, S., Zhang, Y., Zhang, Y., and Lee, H. (2015). Deep visual analogy-making. In *Neural information processing systems*, pages 1252–1260.