

Article

# A Deep Learning Model of Perception in Color-Letter Synesthesia

Joel R. Bock<sup>1</sup>

[1]<sup>1</sup> San Diego, CA, 92120, USA; sauerkraut@gmail.com

**Abstract:** Synesthesia is a psychological phenomenon where sensory signals become mixed. Input to one sensory modality produces an experience in a second, unstimulated modality. In “grapheme-color synesthesia”, viewed letters and numbers evoke mental imagery of colors. The study of this condition has implications for increasing our understanding of brain architecture and function, language, memory and semantics, and the nature of consciousness. In this work, we propose a novel application deep learning to model perception in grapheme-color synesthesia. Achromatic letter images, taken from database of handwritten characters, are used to induce synesthesia. Results show the model learns to accurately create a colored version of the inducing stimulus, according to a statistical distribution from experiments on a sample population of grapheme-color synesthetes. The spontaneous, creative mental imagery characteristic of the synesthetic perceptual experience is reproduced by the model. A model of synesthesia that generates testable predictions on brain activity and behavior is needed to complement large scale data collection efforts in neuroscience, especially when articulating simple descriptions of cause (stimulus) and effect (behavior). The research and modeling approach reported here begins to address this need.

**Keywords:** synesthesia; deep learning network; color perception; generative adversarial network; cognitive modeling; character recognition; GPU computing

## 1. Introduction

### 1.1. Background

Synesthesia is a psychological phenomenon where sensory signals become mixed; input to one sensory modality produces an experience in a second, unstimulated modality [1]. For example, the experience of colors may be induced by seeing or hearing digits, letters or words. In “grapheme-color synesthesia”, viewed letters and numbers evoke mental imagery of colors. These color associations are involuntary, idiosyncratic and highly consistent over time [2]. The study of synesthesia has implications for furthering our understanding of brain architecture and function, as well as creative processes, language acquisition, and learning and memory performance [3]. The study of consciousness itself may benefit from investigation of synesthesia [4].

Due to the significance of synesthesia across disciplines, the scientific literature is rich with studies aiming to account for the physiological origins and behavioral manifestations. Investigations into cognitive and behavioral aspects and neurological substrates are reviewed in [1] and [5].

Cognitive models have been proposed to identify partitions of cognition and perception in synesthesia. In [6], multiple interconnected pathways of form and color analysis under visual or auditory stimuli are modeled. Symbol and color representation domains interact with semantic identification at higher stages of processing. An important contribution of the model in [6] is explicit description of the levels of inducer processing, from preconscious feature analysis to ultimate production of synesthetic assignment of color to the input.

Descriptions of the neurological basis of synesthesia center on two main theories: the synesthetic brain is (1) *cross-activation* between color processing area V4 and proximal visual word form area in the fusiform gyrus ([7]); or (2) *disinhibited feedback* between circuits of bottom-up sensory input and

39 higher-level visual areas [8]. Functional neuroimaging studies dynamically localize activated cortical  
 40 regions in synesthetic perception versus controls [9], and provide valuable insight to validate and  
 41 continue to develop such theories [10].

42 Such neuroscientific investigations are now collecting huge amounts of data, even down to the  
 43 granularity of single neuronal firings. How do cognition and behavior arise from the interactions  
 44 between activated neurons and assemblies under external sensory input? An ongoing challenge is to  
 45 uncover causal relationships between "big neural" and "big behavioral" data [11].

### 46 1.2. Deep learning models

47 Deep learning facilitates machine learning from large scale data. Complicated, abstract  
 48 representations of structure in computer vision, language processing, and many other domains can be  
 49 explored [12], [13].

50 Hinton demonstrated that multilayer generative models could learn the joint distribution  
 51 of handwritten digit images and their labels [14]. These deep belief networks learned latent  
 52 representations of the input in densely-connected hidden layers. Generative models were shown  
 53 capable of (1) learning low-level features in an unsupervised manner, and (2) learning very large  
 54 number of parameters without overfitting [14].

55 Recently, Goodfellow [15] introduced generative adversarial networks (GANs), a general  
 56 framework for training deep learning networks. GANs eliminate the need for difficult probabilistic  
 57 computations when learning hidden layer parameters.

58 The central idea of GANs is to establish competition between two deep network models– the  
 59 *discriminator* ( $D$ ) and the *generator* ( $G$ ).  $G$  is tasked with generating samples  $G(\mathbf{z})$  (drawn from  $p_{\mathbf{z}}(\mathbf{z})$ )  
 60 that appear to  $D$  as having been drawn from the actual distribution  $p_{data}(\mathbf{x})$ .  $D$  must learn to discern  
 61 between real and artificial data. In this numerical game, model parameters are optimized alternatively  
 62 to solve the minimax objective function

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} [\log(D(\mathbf{x}))] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))] \quad (1)$$

63 This formulation has an unique global optimum representing the real data distribution  $p_{data}(\mathbf{x})$   
 64 [15]. This is true even when the prior distribution  $p_{\mathbf{z}}(\mathbf{z})$  is random noise. Extensions of the GAN  
 65 framework condition  $D$  and  $G$  on additional information such as class label [16], [17].

66 Both  $D$  and  $G$  are implemented as multilayer convolutional neural networks, as described in  
 67 Section 2.3.

### 68 1.3. Contribution of present study

69 In this work, the generator network  $G$  is a model for grapheme-color synesthesia. Achromatic  
 70 letter images, taken from database of handwritten characters [18], are used to induce synesthesia.

71  $G$  learns to create a colored version of the inducing stimulus, according to a statistical distribution  
 72 from experiments on grapheme-color synesthetes [19]. The identity of each symbol determines its  
 73 concurrent color [8].

74 The spontaneous, creative mental imagery characteristic of the synesthetic perceptual experience  
 75 is reproduced by the model.

76 The present study contributes an approach to extract essential mechanisms relating sensory input  
 77 and processing to behavioral effects in neuro- and cognitive science.

## 78 2. Methods

### 79 2.1. Handwritten letters database

80 Handwritten letter images were extracted from the EMNIST dataset [18]. The raw images are  
 81 stored as 28X28 pixels, in 8-bit integer format. A modeling sample comprising the EMNIST uppercase

82 letters was constructed using the *By\_Class* subset and annotations. We excluded lowercase letters and  
 83 numeric digits resulting in 220,304 examples. Counts of individual letters in this sample varied from  
 84 2,850 ('K') to >29,000 ('O'); qualitatively, moderate variance in handwritten morphological structure for  
 85 given letters was observed. No balancing of the 26 letter-classes was carried out; the aim here was not  
 86 to develop a discriminative model for classification.<sup>1</sup>

## 87 2.2. Synesthesia color-letter pairs

88 Each grayscale letter image was converted to a 3-channel (*R,G,B*) image using experimental  
 89 statistics of perceived colors in grapheme-color synesthesia as reported by Witthoft *et al.* [19]. The most  
 90 frequently reported letter-color pairings from a large sample of synesthetes ( $n=6,588$ ) were used to  
 91 represent the sample population (c.f. Fig.1 in [19]), recognizing that significant idiosyncratic differences  
 92 in color experienced for a given letter exists between individuals. These aggregate "modal" colors for  
 93 each letter were used to develop the basic examples for generative colorization model learning in the  
 94 current study. These pairings are listed in Table 1.

**Table 1.** Grapheme-color associations of 6,588 synesthetes. The most common color assignment reported for each letter is shown. After [19].

Color	Letters
Red	A, M, R
Blue	B, D, T, W
Green	E, F, G
Yellow	C, L, Y
Orange	H, J, K, N, U
Purple	P, Q, V
White	I, O
Black	X, Z

## 95 2.3. Numerical implementation

96 The conditional GAN model of grapheme-color synesthesia perception was adapted from a  
 97 deep convolutional neural network (CNN) implementation described in [20]. The generator network  
 98 encodes the input image by six successive hidden layers, each outputting a reduced dimensional  
 99 image relative to the input. Six decoding layers follow; these layers invert the compression operations  
 100 by deconvolution. The final output of the generator is an image with synthetic colorization. Each  
 101 compression layer reduces noise while increasing the level of abstraction in representation of the input  
 102 [21]. The discriminator has similar multilayer convolutions, ultimately outputting a reduced image and  
 103 probabilities of the input image being real or synthetic. Additional details of the GAN architecture  
 104 appear in [20].<sup>2</sup>

105 We developed code to download, extract and preprocess EMNIST handwritten letters [18]; colorize  
 106 examples based on synesthesia statistics [19]; train, test, analyze and visualize results of the generative  
 107 learning process. Deep learning experiments were carried out using the TensorFlow software library  
 108 controlled by the Python language API [22]<sup>3</sup>. Models were developed on a CPU/GPU-based system  
 109 (i5-3470 16 GB RAM; NVIDIA GeForce GTX 1050 Ti 4 GB on-board).

110 Transformations and processing for each training image followed the protocol of [20]: (1) convert  
 111 example (*R,G,B*) image (described in Sections 2.1 and 2.2) to CIE Lab color space; (2) present the  
 112 *L* (grayscale) channel to the input of the generative model *G*; (3) output synthetic *a,b* channel data

<sup>1</sup> The EMNIST data can be obtained at: <https://www.nist.gov/itl/iad/image-group/emnist-dataset>.

<sup>2</sup> See [20] Appendix 1. Original image colorization GAN adapted here is based on code from the repository: <https://github.com/sawhney-kartik/gan-colorization>.

<sup>3</sup> All computer code developed in this study is available to researchers at: [boris@github.com/TBD](mailto:boris@github.com/TBD).

113 from  $G$ , and reconstruct a full 3-channel ( $L,a,b$ ) color image; (4) present real and artificial images to  
 114 discriminator  $D$ ; (5) evaluate objective function (Equation 1); (6) backpropagate errors through both  
 115 networks, updating weights using stochastic gradient descent.

116 Generative model training was performed for three epochs on a random sample of 125,000 images  
 117 (total 375,000 iterations) from the EMNIST dataset. Periodically, generated images were converted  
 118 back to ( $R,G,B$ ) color space and stored to disk for human observation and post-analysis.

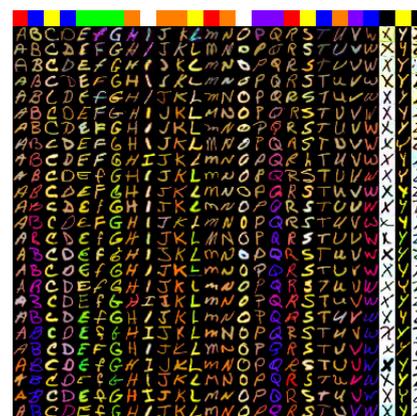
### 119 3. Results

120 Results of synesthetic letter colorization by the generative network are presented in Figure 1.  
 121 The figure panels display (a) typical inputs and (b) corresponding outputs in the very early phases  
 122 of model training ( $< 3,000$  iterations). Each panel contains samples for one handwritten letter per  
 123 column. Iteration count progresses from top to bottom. The top-most row on the right-hand side  
 124 shows the true modal color-letter pairings from the experimental distribution reported in [19] (Table 1).  
 125 Letters are presented in random order and are unbalanced; the example count seen for each letter is  
 126 therefore not uniform.

127 At this stage of learning, the generative network is not producing realistic colored images matching  
 128 the true distribution. Several letters begin to appear to align with their synesthetic concurrents, but  
 129 most others miss the mark. The letters often perceived as either black or white in color-grapheme  
 130 synesthesia (I,O,X,Z) are already reproduced fairly well.

131  
132  
133  
134  
135

(a) Early generator input.



(b) Early generator color induction.

**Figure 1. Synesthetic letter colorization by the evolving generative network.** Early training: 1st epoch;  $< 3,000$  iterations. (a) generator input; (b) generator output. One letter per column. Iteration count progresses from top to bottom.

131 Following additional optimization of the GAN, more realistic results are observed. Consider  
 132 the images shown in Figure 2. After three epochs of training, the synesthesia model generates  
 133 colored letters with high accuracy. The colors produced are nearly indistinguishable from the actual  
 134 distribution. Accuracy of color reproduction in the 26X26 grid is over 99%; very few instances (G: row  
 135 7; F: row 18; K: row 2; O: row 15; Q: row 13; Y: row 9) are incorrectly colored.

136 The results exhibited in Figure 2 are typical of those observed in out-of-sample tests on ~95,000  
 137 additional handwritten letter examples. Once the generator learns the true distribution of color-letter  
 138 associations, the concurrent response remains consistent under additional stimuli.



(a) Trained generator input.

(b) Trained generator color induction.

**Figure 2. Synesthetic letter colorization by the trained generative network.** Late training: 3rd epoch, ~375,000 iterations. (a) generator input; (b) generator output. One letter per column. Iteration count progresses from top to bottom.

#### 139 4. Discussion

140 We applied a generative deep neural network to model perception in grapheme-color synesthesia.  
 141 Grayscale letters are used to stimulate the model, which colorizes each letter. Training data were taken  
 142 from experiments on a large sample of synesthetes [19], combined with a database of handwritten  
 143 letters [18].

144 Automatic and consistent response to characters or digits are fundamental to synesthesia, even  
 145 when the stimulus is non-physical (i.e., conceived in the mind's eye) [23]. Recognition of the identity of  
 146 a symbol determines its concurrent color [8], [2]. In the present model, the generative network hidden  
 147 layer weights encode information on pixel intensities describing the structure of each letter (and its  
 148 identity) when perceived in two dimensions. We suggest that a similar mechanism may regulate the  
 149 process of letter identification and association with color in a synesthete's perceptual experience.

150 A model of synesthesia that generates testable predictions on brain activity and behavior is needed  
 151 [2]. With the collection of "big data" in neuroscience, comes an increased complexity of analysis and  
 152 the distillation of conclusions, especially when articulating simple descriptions of cause (stimulus) and  
 153 effect (behavior) [11]. The research and modeling approach reported here begins to address this need.

154 Three particularly ambitious areas in which to apply and extend this research are summarized  
 155 below.

- 156 1. *Functional brain imaging.* Studies aimed to identify and localize structure and function within  
 157 the synesthesia experience [5]. Experimental data from functional magnetic resonance imaging  
 158 (fMRI) or other modalities could be used to build a complementary deep learning model with  
 159 more explicit mapping of layers to physiologic modular components than used in the current  
 160 study. Such a model could provide additional insights to refine the competing hypotheses of  
 161 cross-activation (increased linkage between proximal regions) [7] or disinhibited feedback from  
 162 higher-level cortical areas [8] in synesthesia.

- 163 2. *Language learning and memory.* Studying synesthesia may advance our understanding of human  
164 perception and information arrangement [6]. One theory proposed in [3] advances the idea  
165 that graphemecolor synesthesia develops in part by children to learn category structures; a  
166 fundamental task in literacy development is to recognize and discriminate between letters.  
167 Therefore synesthesia might arise as an aid to memory. More generally, the ability to discern  
168 statistical regularities of printed letters or learn complex rules for letter combinations would assist  
169 learning at subsequent stages of literacy development [3]. In [24], small sample experiments  
170 suggested that additional sensory dimensions in synesthesia aid in memory tasks when compared  
171 to controls. We submit that the generative modeling approach of the current work may be useful  
172 to develop and test hypotheses in studies of language acquisition, memory and semantics.
- 173 3. *Consciousness studies.* Synesthesia can give insight into the neural correlates of consciousness,  
174 through interaction between sensory inputs and their mediation by semantics in the induction  
175 of phenomenal subjective experience [4]. Connecting neural activations with subjective  
176 aspects of consciousness (perception of shape, color movement of an object) is potentially  
177 achievable following a systematic experimental approach [25]. In deep learning, understanding  
178 representations within deep layers is easy at the layer level; at the level of individual neurons,  
179 such understanding is much more difficult [26]. Extensions of the deep learning model reported  
180 here may help to advance toward these formidable objectives.

181 **Acknowledgments:** The author thanks Tom Zingale, who provided the GPU-based computer used to conduct  
182 this research.

183 **Conflicts of Interest:** The author declares no conflict of interest.

## 184 Abbreviations

185 The following abbreviations are used in this manuscript:

186	GAN	Generative adversarial network
	GPU	Graphics processing unit
187	CPU	Central processing unit
	CNN	Convolutional neural network
	fMRI	Functional magnetic resonance imaging

## 188 References

- 189 1. Hubbard, E.M.; Ramachandran, V. Neurocognitive Mechanisms of Synesthesia. *Neuron* **2005**, *48*, 509–520.
- 190 2. Mattingley, J. Attention, Automaticity, and Awareness in Synesthesia. *Annals of the New York Academy of*  
191 *Sciences* **2009**, *1156*, 141–67.
- 192 3. Watson, M.; Akins, K.; Spiker, C.; Crawford, L.; Enns, J. Synesthesia and learning: a critical review and  
193 novel theory. *Frontiers in Human Neuroscience* **2014**, *8*, 98.
- 194 4. van Leeuwen, T.M.; Singer, W.; Nikolić, D. The Merit of Synesthesia for Consciousness Research. *Frontiers*  
195 *in Psychology* **2015**, *6*, 1850.
- 196 5. Rouw, R.; Scholte, H.; Colizoli, O. Brain Areas Involved in Synaesthesia: A Review. *Journal of*  
197 *Neuropsychology* **2011**, *5*, 214–42.
- 198 6. Rich, A.; Mattingley, J. Anomalous perception in synaesthesia: A cognitive neuroscience perspective.  
199 *Nature Reviews Neuroscience* **2002**, *3*, 43–52.
- 200 7. Ramachandran, V.; Hubbard, E. Psychophysical investigations into the neural basis of synaesthesia.  
201 *Proceedings of the Royal Society of London B: Biological Sciences* **2001**, *268*, 979–983.
- 202 8. Grossenbacher, P.; Lovelace, C. Mechanisms of synesthesia: cognitive and physiological constraints.  
203 *TRENDS in Cognitive Sciences* **2001**, *5*, 36–41.
- 204 9. Jäncke, L.; Beeli, G.; Eulig, C.; Hänggi, J. The neuroanatomy of grapheme-color synesthesia. *European*  
205 *Journal of Neuroscience* **2009**, *29*, 1287–93.

- 206 10. Brang, D.; Hubbard, E.; Coulson, S.; Huang, M.; Ramachandran, V. Magnetoencephalography reveals early  
207 activation of V4 in grapheme-color synesthesia. *NeuroImage* **2010**, *53*, 268–274.
- 208 11. Sejnowski, T.; Churchland, P.; Anthony Movshon, J. Putting big data to good use in neuroscience. *Nature*  
209 *Neuroscience* **2014**, *17*, 1440–1.
- 210 12. Bengio, Y. Learning deep architectures for AI. *Foundations and Trends in Machine Learning* **2009**, *2*, 1–127.
- 211 13. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–44.
- 212 14. Hinton, G.E.; Osindero, S.; Teh, Y.W. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*  
213 **2006**, *18*, 1527–1554.
- 214 15. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio,  
215 Y. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems 27*; Ghahramani,  
216 Z.; Welling, M.; Cortes, C.; Lawrence, N.D.; Weinberger, K.Q., Eds.; Curran Associates, Inc., 2014; pp.  
217 2672–2680.
- 218 16. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *CoRR* **2014**, *abs/1411.1784*.
- 219 17. Gauthier, J. Conditional generative adversarial networks for convolutional face generation. Technical  
220 report, Stanford University, 2015.
- 221 18. Cohen, G.; Afshar, S.; Tapson, J.; van Schaik, A. EMNIST: an extension of MNIST to handwritten letters.  
222 *CoRR* **2017**, *abs/1702.05373*, [[1702.05373](#)].
- 223 19. Witthoft, N.; Winawer, J.; Eagleman, D.M. Prevalence of Learned Grapheme-Color Pairings in a Large  
224 Online Sample of Synesthetes. *PLoS ONE* **2015**, *10*.
- 225 20. Agrawal, M.; Sawhney, K. Exploring Convolutional Neural Networks for Automatic Image Colorization.  
226 Technical report, Stanford University, 2017.
- 227 21. Shwartz-Ziv, R.; Tishby, N. Opening the Black Box of Deep Neural Networks via Information. *CoRR* **2017**,  
228 *abs/1703.00810*, [[1703.00810](#)].
- 229 22. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard,  
230 M.; Kudlur, M.; Levenberg, J.; Monga, R.; Moore, S.; Murray, D.G.; Steiner, B.; Tucker, P.A.; Vasudevan, V.;  
231 Warden, P.; Wicke, M.; Yu, Y.; Zhang, X. TensorFlow: A system for large-scale machine learning. *CoRR*  
232 **2016**, *abs/1605.08695*, [[1605.08695](#)].
- 233 23. Dixon, M.J.; Smilek, D.; Cudahy, C.; Merikle, P.M. Five plus two equals yellow. *Nature* **2000**, *406*, 365.
- 234 24. Mills, C.; Innis, J.; Westendorf, T.; Owsianiecki, L.; McDonald, A. Effect of a Synesthete's Photisms on  
235 Name Recall. *Cortex* **2006**, *42*, 155–63.
- 236 25. Crick, F.; Koch, C. A Framework for Consciousness. *Nature Neuroscience* **2003**, *6*, 119–26.
- 237 26. Hinton, G. Learning multiple layers of representation. *Trends in Cognitive Sciences* **2007**, *11*, 428–434.