# Data-driven generative models for perception, dreaming, and imagining

University of Minnesota, Spring Semester, 2018

## **Topics in Computational Vision

**Psy 8036 (Kersten)
Psy 5993 Section 034 (Schrater)

http://courses.kersten.org

## https://ay17.moodle.umn.edu/course/view.php?id=8619#section-11

**Instructors:**
Dan Kersten (kersten@umn.edu)
Paul Schrater (schrater@umn.edu)

**Summary**

It has been proposed that perception is fundamentally a process of "analysis-by-synthesis" in which the sensory input is analyzed bottom-up, with perceptual interpretations tested and refined by top-down predictions of the input, through synthesis. However, while the computational and neural study of the analysis component is well-developed, less is known about the principles and mechanisms that underly synthesis. This seminar will explore recent advances using "deep" learning algorithms to discover hierarchical statistical regularities in large datasets of natural patterns, and the relevance of the learning results to models of human perception and recognition. These algorithms also provide the basis for the stochastic synthesis of novel, yet familiar patterns, which raises the question of whether the human experiences of dreams and hallucinations, and the ability to imagine, reflect the same statistical regularities that are discoverable using machine learning. The class format will include short

introductory lectures by the instructors, and weekly student presentations of current literature. The short lectures will provide historical context as well as tutorials on machine learning (e.g. TensorFlow for neural network simulations).

**Meeting time**: First meeting Tuesday, Jan 16$^{th}$, 3:00 pm.
**Place: Elliott N227**

Students can sign up for either Topics in Computational Vision Psy 8036 (Kersten) or Psy 5993 Section 034 (Schrater) .

**Background**

There is a long history of theories of perception in which the brain “explains” sensory input in terms of external, behaviorally relevant causes. A current hypothesis is that this process is implemented in part by cortical feedback mechanisms that synthesize predictions of early data representations in order to test how well the brain's current interpretation of the world corresponds with the sensory data. In this view, perception involves a cycle in which the incoming data triggers a set of explanations, i.e. hypotheses, which are used to measure how far the expected sensory input differs from the actual input. From a computational perspective, such generative models of perceptual inference have a number of advantages over strictly bottom-up inference. A generative model can incorporate measures of "goodness-of-fit" to decide whether to accept or reject an interpretation--some explanations are better than others. Discrepancies between sensory data and predictions may also be used to direct attentional resources and signal whether more complex combinations of hypotheses are needed. Further, with sufficient structure, a generative model could provide the basis for the perceptual interpretation of sensory input outside the range of past experience.

While computational theories for bottom-up neural mechanisms for perception have received considerable scientific attention, much less is known about top-down mechanisms. This seminar will explore the idea that the brain has hierarchically structured mechanisms that can synthesize patterns of input representations with the following constraints: 1) the mechanisms build on inductive structural biases that are innate; 2) the mechanisms reflect the statistical regularities induced by the physical causes of sensory experience, i.e. they are "data-driven"; 3) the need for cognitive processes to access semantic, perceptual content over levels of abstraction. Assumptions 1) and 2) constrain the class of generative models to be "data-driven", i.e. models that can be learned from sensory data.

Recent computational methods for data-driven pattern synthesis (e.g. VAE, InfoGAN, Adversarial Bayes, StackGAN) will be covered in this seminar.  We will also explore the proposal that the same circuitry that may underly feedback in perception is used during imagery, dreams, and hallucinations.

## *Tentative Syllabus*

| **Week** | **Topics** | **Background material** | **Discussion topics and papers** |
|---|---|---|---|
| 1: Jan 16 | Background Models of perception | Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? Trends in Cognitive Sciences, 10(7), 301–308\. | |
| 2: Jan 23 | Overview of machine learning | Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. Cognitive Science, 9(1), 147–169. | |
| 3: Jan 30 | Shallow image models, textures | Zhu, S. C., Wu, Y., & Mumford, D. (1998). Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling. International Journal of Computer Vision, 27(2), 107–126\. McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. Nature Publishing Group, 16(4), 493–498\. | |
| 4: Feb 6 | Hierarchical image models, deep learning | Zhu, S.-C., & Mumford, D. (2006). Quest for a stochastic grammar of images. Foundations and Trends® in Computer Graphics and Vision, 2(4), 259–362\. | Topic preview: Visual imagery |
| 5: Feb 13 | Hierarchical image models, deep learning | | Topic preview: Auditory imagery |
| 6: Feb 20 | Hierarchical image models | | Topic preview: Hypnagogic |

# *Sample Readings (under construction)*

## Background

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. Cognitive Science, 9(1), 147–169.
Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. Neuron, 76(4), 695–711.
Berkes, P., Orban, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. Science, 331(6013), 83–87.
Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz Machine. Neural Computation, 7(5), 889–904.
Ouden, den, H. E. M. (2012). How prediction errors shape perception, attention, and motivation, 1–12.
Orban, G., Pietro Berkes, Fiser, J., & Lengyel, M. (2016). Neural Variability and Sampling-Based Probabilistic Representations in the Visual Cortex. Neuron, 92(2), 530–543.
MacKay, D. M. (1956). Towards an information-flow model of human behaviour. British Journal of Psychology (London, England : 1953), 47(1), 30–43.
Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. Science, 350(6266), 1332–1338. http://doi.org/10.1126/science.aab3050
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. http://doi.org/10.1038/nature14539
McDermott, Josh H., and Andrew J. Oxenham. "Spectral Completion of Partially Masked Sounds." Proceedings of the National Academy of Sciences 105, no. 15 (2008): 5939–5944.Mumford, D. (1992). On the computational architecture of the neocortex. Biological Cybernetics, 66(3), 241–251.
Mumford, D. (1994). Pattern theory: a unifying perspective, 187–224.
Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects.

Nature Neuroscience, 2, 79–87.
Tu, Z., Chen, X., Yuille, A. L., & Zhu, S.-C. (2005). Image parsing: Unifying segmentation, detection, and recognition. International Journal of Computer Vision, 63(2), 113–140.
Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? Trends in Cognitive Sciences, 10(7), 301–308.
Richards, W. (1971). The Fortification Illusions of Migraines, Scientific American, 1–10.
Zhu, S.-C., & Mumford, D. (2006). Quest for a stochastic grammar of images. Foundations and Trends® in Computer Graphics and Vision, 2(4), 259–362. http://doi.org/10.1561/0600000018

## **Shallow generative models**: **Texture synthesis**

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. Nature Publishing Group, 14(9), 1195–1201. http://doi.org/10.1038/nn.2889
McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. Nature Publishing Group, 16(4), 493–498.
McDermott, J. H., & Simoncelli, E. P. (2011). Sound Texture Perception via Statistics of the Auditory Periphery: Evidence from Sound Synthesis. Neuron, 71(5), 926–940.
Zhu, S. C., Wu, Y., & Mumford, D. (1998). Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling. International Journal of Computer Vision, 27(2), 107–126.

## Hierarchical (deep) data-driven generative models

Chen, X., Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets, 2172–2180.
Goodfellow, I. (2016, December 31). NIPS 2016 Tutorial: Generative Adversarial Networks.
Kulkarni, T. D., Whitney, W. F., Kohli, P., & Tenenbaum, J. (2015). Deep Convolutional Inverse Graphics Network, 2539–2547.
Rock, J., Issaranon, T., Deshpande, A., & Forsyth, D. (2016, December 5). Authoring image decompositions with generative models.

Varol, G., Romero, J., Martin, X., Mahmood, N., Black, M. J., Laptev, I., & Schmid, C. (2017, January 5). Learning from Synthetic Humans.
Xie, J., Zhu, S.-C., & Wu, Y. N. (2016, June 3). Synthesizing Dynamic Patterns by Spatial-Temporal Generative ConvNet.
Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., & Lipson, H. (2015, June 22). Understanding Neural Networks Through Deep Visualization.
Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. (2016, December 10). StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks.

## Hypnagogic imagery

Gurstelle, E. B., & de Oliveira, J. L. (2004). Daytime parahypnagogia: a state of consciousness that occurs when we almost fall asleep. Medical Hypotheses, 62(2), 166–168. http://doi.org/10.1016/S0306-9877(03)00306-2
Holmes, E. A., James, E. L., Coode-Bate, T., & Deeprose, C. (2009). Can Playing the Computer Game "Tetris" Reduce the Build-Up of Flashbacks for Trauma? A Proposal from Cognitive Science. PLoS ONE, 4(1), e4153. http://doi.org/10.1371/journal.pone.0004153.t004
Nielsen, T. A. (1995). Describing and modeling hypnagogic imagery using a systematic self-observation procedure. Dreaming, 5(2), 75–94. http://doi.org/10.1037/h0094426
Nielsen, T. A. (2016). A Self-Observational Study of Spontaneous Hypnagogic Imagery Using the Upright Napping Procedure. Imagination, Cognition and Personality, 11(4), 353–366. http://doi.org/10.2190/3LVV-L5GY-UR5V-N0TG
*Schacter, D. L. (1976). The hypnagogic state: a critical review of the literature. Psychological Bulletin.
Stickgold, R. (2000). Replaying the Game: Hypnagogic Images in Normals and Amnesics. Science, 290(5490), 350–353. http://doi.org/10.1126/science.290.5490.350

## Dreams

Band, J. C. Z. F. A., 2016. (n.d.). Animal "Hypnosis" and Waking Nightmares. Anomalistik.De
Crick, F., G. Mitchison., 1983. The function of dream sleep. Nature. Springer

*Hobson, J. A., & Mccarley, R. W. (197.). The brain as a dream state generator: an activation-synthesis hypothesis of the dream process. The American Journal of Psychiatry.
Dresler, M., Koch, S. P., Wehrle, R., Spoormaker, V. I., Holsboer, F., Steiger, A., et al. (2011). Dreamed Movement Elicits Activation in the Sensorimotor Cortex. Current Biology : CB.
Stickgold, R., Hobson, J. A., Fosse, R., & Fosse, M. (2001). Sleep, Learning, and Dreams: Off-line Memory Reprocessing. Science, 294(5544), 1052–1057. http://doi.org/10.1126/science.1063530
Stickgold, R. (2005). Sleep-dependent memory consolidation. Nature, 437(7063), 1272–1278. http://doi.org/10.1038/nature04286
Studies, J. H. J. O. C., 2014. (n.d.). Consciousness, dreams, and inference: the cartesian theatre revisited. Ingentaconnect.com

## Hallucinations

Bressloff, P. C., Cowan, J. D., Golubitsky, M., Thomas, P. J., & Wiener, M. C. (2002). What geometric visual hallucinations tell us about the visual cortex. Neural Computation, 14(3), 473–491. http://doi.org/10.1162/089976602317250861
Cummings, J. L., & Miller, B. L. (1987). Visual hallucinations. Clinical occurrence and use in differential diagnosis. The Western Journal of Medicine, 146(1), 46–51.
*Ermentrout, G. B., & Cowan, J. D. (1979). A mathematical theory of visual hallucination patterns. Biological Cybernetics, 34(3), 137–150. http://doi.org/10.1007/BF00336965
Merabet, L. B., Maguire, D., Warde, A., Alterescu, K., Stickgold, R., & Pascual-Leone, A. (2004). Visual hallucinations during prolonged blindfolding in sighted subjects. Journal of Neuro-Ophthalmology, 24(2), 109–113.
Howard, R. J., Brammer, M. J., David, A., Woodruff, P., & Williams, S. (1998). The anatomy of conscious vision: an fMRI study of visual hallucinations. Nature neuroscience, 1(8), 738-742.Seriès, P., Reichert, D. P., & Storkey, A. J. (2010). Hallucinations in Charles Bonnet Syndrome Induced by Homeostasis: a Deep Boltzmann Machine Model, 2020–2028.
Silverstein, S. M. (2016). Visual Perception Disturbances in Schizophrenia: A Unified Model. In The Neuropsychopathology of Schizophrenia: Molecules,

Brain Systems, Motivation, and Cognition (3rd ed., Vol. 63, pp. 77–132). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-30596-7_4
Kumar, S., Sedley, W., Barnes, G. R., Teki, S., Friston, K. J., & Griffiths, T. D. (2014). A brain basis for musical hallucinations. Cortex, 52(C), 86–97. http://doi.org/10.1016/j.cortex.2013.12.002

## Imagery and imagination

Chetverikov, A., & Kristjánsson, Á. (2016). On the joys of perceiving: Affect as feedback for perceptual predictions. Actpsy, 169(C), 1–10. http://doi.org/10.1016/j.actpsy.2016.05.005
Dijkstra, N., Zeidman, P., Ondobaka, S., Gerven, M. A. J., & Friston, K. (2017). Distinct Top-down and Bottom-up Brain Connectivity During Visual Perception and Imagery. Scientific Reports, 1–9. http://doi.org/10.1038/s41598-017-05888-8
Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., & Ondobaka, S. (2017). Active Inference, Curiosity and Insight. Neural Computation, 29(10), 2633–2683. http://doi.org/10.1162/neco_a_00999
Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? Psychological Bulletin, 129(5), 723–746. http://doi.org/10.1037/0033-2909.129.5.723
Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., et al. (1993). Visual Mental Imagery Activates Topographically Organized Visual Cortex: PET Investigations. Journal of Cognitive Neuroscience, 5(3), 263–287. http://doi.org/10.1162/jocn.1993.5.3.263
Kosslyn, S., & Ganis, G. (2000). Neural foundations of imagery. Nature Reviews ….
Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental Imagery: Functional Mechanisms and Clinical Applications. Trends in Cognitive Sciences, 19(10), 590–602. http://doi.org/10.1016/j.tics.2015.08.003
Riecke, L., A. J. van Opstal, R. Goebel, and E. Formisano. “Hearing Illusory Sounds in Noise: Sensory-Perceptual Transformations in Primary Auditory Cortex.” Journal of Neuroscience 27, no. 46 (November 14, 2007): 12684–89. https://doi.org/10.1523/JNEUROSCI.2713-07.2007.
Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., &

Szpunar, K. K. (2012). The Future of Memory: Remembering, Imagining, and the Brain. Neuron, 76(4), 677–694. http://doi.org/10.1016/j.neuron.2012.11.001
Zatorre, R. J., & Halpern, A. R. (2005). Mental Concerts: Musical Imagery and Auditory Cortex. Neuron, 47(1), 9–12. http://doi.org/10.1016/j.neuron.2005.06.013

## Imagery and memory

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex. Curbio, 23(15), 1427–1431. http://doi.org/10.1016/j.cub.2013.05.065
Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The Distributed Nature of Working Memory. Trends in Cognitive Sciences, 1–15. http://doi.org/10.1016/j.tics.2016.12.007
Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. NeuroImage, 105(C), 215–228. http://doi.org/10.1016/j.neuroimage.2014.10.018
Self, M. W., van Kerkoerle, T., & Roelfsema, P. R. (2016). Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. Nature Communications, 8, 1–12. http://doi.org/10.1038/ncomms13804
Stickgold, R. (2005). Sleep-dependent memory consolidation. Nature, 437(7063), 1272–1278. http://doi.org/10.1038/nature04286