ISM Annotated Bibliography #2

Thomas Ray

Independent Study Mentorship

September 18, 2022

Gardner, M., Grus, J., Neumann, M., Tafjord, O., Dasigi, P., Liu, N., Peters, M., Schmitz, M., & Zettlemoyer, L. (2018, May 31). Allennlp: A deep semantic natural language processing platform. arXiv.org. Retrieved September 23, 2022, from https://arxiv.org/abs/1803.07640

In this study, AllenNLP, a research platform for deep learning approaches to natural language comprehension, is introduced. AllenNLP is made to help researchers who wish to rapidly and simply create unique language understanding models. It is built on top of PyTorch, enabling dynamic computation graphs, and offers a configurable data API that manages intelligent batching and padding, high-level abstractions for typical text-related operations, and a modular and extendable experiment structure that facilitates doing sound research. Additionally, it contains examples of reference implementations of superior strategies for both applications involving language understanding and fundamental semantic issues. AllenNLP is an ongoing open-source effort maintained by engineers and researchers at the Allen Institute for Artificial Intelligence.

Ghose, S., & Prevost, J. J. (2020). AutoFoley: Artificial Synthesis of Synchronized Sound Tracks for Silent Videos with Deep Learning. IEEE Transactions on Multimedia, 1–1. doi:10.1109/tmm.2020.3005033

In this paper, Dr. Ghose and Dr. Prevost present Autofoley, an deep learning tool that can create sound effects for videos. Foley artists usually make these sound effects but the Professors theorize that Autofoley could replace them, significantly cutting down production costs for various media. The paper proposes 2 methods in order to create Autofoley. The first method consists of utilizing an RNN LSTM specifically

ResNet-FSLSTM which utilizes interpolation to obtain more information from a video. The second method still utilizes interpolation and ResNet but now uses TRN instead of LSTM which means that we only learn from the frames available in the video. Both models have their advantages with the first model being better for fast-moving clips while the second model is better for early recognition of an activity using limited video frames. Once we have identified what's happening in the video we can use the ISTFT method to generate sound.

Herrera-Aguilar, J. L., Larralde, H., & Aldana, M. (2012). Pattern Recognition in Neural Networks with Competing Dynamics: Coexistence of Fixed-Point and Cyclic Attractors. PLoS ONE, 7(8), e42348. https://link.gale.com/apps/doc/A543306337/SCIC?u=j043905009&sid=bookmark-SCIC&xid=76d19b83

This study investigates the characteristics of the dynamical phase shift that takes place in neural network models where sequential pattern recognition and associative memory are in conflict. This competition is carried out via a weighted combination of the symmetric and asymmetric synaptic matrix components. Using a generating functional formalism, we identify the regions of high and weak pattern correlations, the spin-glass solutions, and the order-disorder transition between these regions in order to determine the structure of the parameter space at non-zero temperature and near saturation (i.e., when the number of stored patterns scales with the size of the network). This investigation shows that smooth transitions between highly correlated regions and fictitious states occur when associative memory is dominating. The transitions, however, are usually discontinuous when sequential pattern recognition dominates associative memory. Additionally, there is

a discontinuous transition between associative memory and sequential pattern recognition when the symmetric and asymmetric sections of the synaptic matrix are specified in terms of the same set of patterns. In contrast, the network may perform both associative memory and sequential pattern recognition for a broad range of parameter values when the symmetric and asymmetric components of the synaptic matrix are specified in terms of distinct sets of patterns.

Liu, N., Li, S., & Du, Y. (n.d.). Compositional visual generation with composable diffusion ... - arxiv. Retrieved September 23, 2022, from https://arxiv.org/pdf/2206. 01714.pdf

Given natural language descriptions, large text-guided diffusion models, like DALLE-2, may produce breathtaking photorealistic visuals. Although these models are quite adaptable, they have trouble comprehending how some ideas are composed, such as when relations between things or the qualities of several objects are mixed up. In this study, a new technique for compositional generation utilizing diffusion models is provided. A group of diffusion models is combined to create a picture, with each model simulating a different aspect of the image. To do this, the researchers interpret diffusion models as energy-based models that allow for the explicit combination of the data distributions specified by the energy functions. The suggested approach can create test-time scenarios that are far more complex than those observed during training, including language descriptions, object relations, human face features, and even generalizing to novel combinations that are infrequently encountered in everyday life. They continue by demonstrating how their method can be used to create pre-trained text-guided diffusion models and produce photorealistic pictures that contain all the features mentioned in the input descriptions, including the binding of several object properties that have proven

challenging for DALLE-2. These outcomes demonstrate how well the suggested approach works to encourage organized generalization for visual creation.

Turakhia, D. G., Allen, H. M., DesPortes, K., & University, N. Y. (2021, June 1). Fabo: Integrating fabrication with a player's gameplay in existing Digital Games: Creativity and Cognition. ACM Conferences. Retrieved September 23, 2022, from https://dl.acm.org/doi/10.1145/3450741.3465239

Making things out of a player's gameplay, such mementos of expensive game goods or unique game controllers fashioned from game objects, broadens the range of ways to interact with digital games. Currently, researchers build these integrated fabrication games from scratch, which takes time and ignores the opportunity to include fabrication with the numerous existing games. However, without access to the source files, integrating fabrication with the real-time gameplay of existing games is difficult. The researchers provide a framework that employs on-screen visual information to combine manufacturing with current digital games in order to overcome this difficulty. The authors created the FabO toolkit to put this framework into practice. Designers use the FabO designer interface to select gameplay moments for fabrication and tag the associated on-screen visual cues. Players use the FabO player interface, which tracks gameplay, recognizes these cues, and automatically generates the fabrication files for the game objects. The findings of our two user tests demonstrate that FabO helped integrate fabrication with a variety of games while enhancing player experience.