Contents

1 Vision

1.1 Scene Understanding

1.1.1 Topic Description

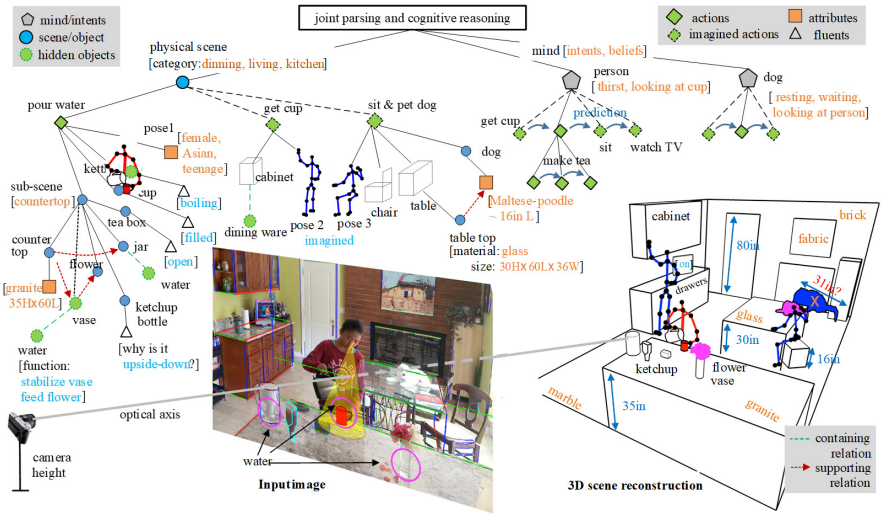

Figure 1: An example of in-depth understanding of a scene or event through joint parsing and cognitive reasoning (Dark, Beyond Deep).

1.1.2 Papers

CVPR 2019/2020/2021 workshops: 3D Scene Understanding for Vision, Graphics, and Robotics.

1. Holistic 3D Scene Parsing and Reconstruction from a Single RGB Image
   Propose a computational framework to parse and reconstruct the 3D configuration of an indoor scene from a single RGB image in an analysis-by-synthesis fasion using a stochastic grammar model.

2. Cooperative Holistic Scene Understanding: Unifying 3D Object, Layout, and Camera Pose Estimation
   Propose an end-to-end model that simultaneously solves tasks of 3D object detection, 3D layout estimation and camera pose estimation in real-time given only a single RGB image.

3. Holistic++ Scene Understanding: Single-view 3D Holistic Scene Parsing and Human Pose Estimation with Human-Object Interaction and Physical Commonsense
   Propose a new 3D holistic++ scene understanding problem, which jointly tackles two tasks from a single-view image: (i) holistic scene parsing and reconstruction- and (ii) 3D human pose estimation. We incorporate the human-object interaction (HOI) and physical commonsense to tackle this problem.

4. Manhattan Room Layout Reconstruction from a Single 360° image: A Comparative Study of State-of-the-art Methods
   Summarize and describe the common framework for predicting layouts from 360° panoramas , the variants, and the impact of the design decisions.

5. SMPL: a Skinned Multi-person Linear Model
   A learned model of human body shape and pose-dependent shape variation.

6. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation
   A unified architecture for applications ranging from object classification, part segmentation, to scene semantic parsing.

7. Understanding Tools: Task-Oriented Object Modeling, Learning and Recognition
   A framework for task-oriented modeling, learning and recognition which aims at understanding the underlying functions, physics and causality in using objects as "tools".

8. Beyond Point Clouds: Scene Understanding by Reasoning Geometry and Physics
   Scene understanding by reasoning physics.

9. Holistic 3D Scene Understanding from a Single Image with Implicit Representation
   Present a new pipeline for holistic 3D scene understanding from a single image, which could predict object shapes, object poses, and scene layout.

10. iGibson: Interactive Simulation of Large Scale Virtualized Realistic Scenes for Robot Learning
iGibson is a simulation environment providing fast visual rendering and physics simulation based on Bullet.

11. BEHAVIOR: Benchmark for Everyday Household Activities in Virtual, Interactive, and Ecological Environments
BEHAVIOR, a benchmark for embodied AI with 100 activities in simulation, spanning a range of everyday household chores such as cleaning, maintenance, and food preparation.

12. Habitat: A Platform for Embodied AI Research
Habitat–a platform for research in embodied artificial intelligence (AI)–enables training embodied agents (virtual robots) in highly efficient photorealistic 3D simulation.

1.1.3 Datasets

1. NYU Depth Dataset V2 (ECCV 2012)
The NYU-Depth V2 data set is comprised of video sequences from a variety of indoor scenes as recorded by both the RGB and Depth cameras from the Microsoft Kinect.

2. SUN RGB-D (CVPR 2015)
Present an RGB-D benchmark suite for the goal of advancing the state-of-the-art in all major scene understanding tasks. The whole dataset is densely annotated and includes 146,617 2D polygons and 58,657 3D bounding boxes with accurate object orientations, as well as a 3D room layout and category for scenes.

3. ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes (CVPR 2017)
ScanNet is an RGB-D video dataset containing 2.5 million views in more than 1500 scans, annotated with 3D camera poses, surface reconstructions, and instance-level semantic segmentations.

4. HoliCity: A City-Scale Data Platform for Learning Holistic 3D Structures
Present HoliCity, a city-scale 3D dataset with rich structural information. Currently, this dataset has 6,300 real-world panoramas of resolution 13312 $\times$ 6656 that are accurately aligned with the CAD model of downtown London with an area of more than 20 km^2.

5. SUNCG: Semantic Scene Completion from a Single Depth Image (CVPR 2017)
A manually created large-scale dataset of synthetic 3D scenes with dense volumetric annotations.

6. Structured3D: A Large Photo-realistic Dataset for Structured 3D Modeling (ECCV 2020)
Structured3D is a large-scale photo-realistic dataset containing 3.5K house designs created by professional designers with a variety of ground truth 3D structure annotations and generate photo-realistic 2D images.

7. 3D-FRONT: 3D Furnished Rooms with layOuts and semaNTics
Introduce 3D-FRONT, a new, large-scale, and comprehensive repository of synthetic indoor scenes highlighted by professionally designed layouts and a large number of rooms populated by high-quality textured 3D models with style compatibility.

8. OpenRooms: An Open Framework for Photorealistic Indoor Scene Datasets (CVPR 2021)
Enhanced ScanNet from a novel framework for creating large-scale photorealistic datasets of indoor scenes, with ground truth geometry, material, lighting and semantics.

1.2 Ego-centric Vision

1.2.1 Topic Description

Study the topics and challenges in ego-centric vision.

> Perceiving the environment includes the *ego* as part of the total process. In order to localize any object, there must be a point of reference. An impression of "there" implies an impression of "here," and neither could exist without the other.
>
> — James Jerome Gibson

1.2.2 Papers

1. LEMMA: A Multi-view Dataset for Learning Multi-agent Multi-task Activities
   Introduces the LEMMA dataset to provide a single home to address these missing dimensions with meticulously designed settings, wherein the number of tasks and agents varies to highlight different learning objectives..

2. Scaling Egocentric Vision: The EPIC-KITCHENS Dataset
   The extended largest dataset in first-person (egocentric) vision; multi-faceted, audio-visual, non-scripted recordings in native environments - i.e. the wearers' homes, capturing all daily activities in the kitchen over multiple days.
   Challenges: EPIC-KITCHENS Workshop.

3. First-Person Activity Forecasting from Video with Online Inverse Reinforcement Learning
   Targets at addressing the problem of incrementally modeling and forecasting long-term goals of a first-person camera wearer: what the user will do, where they will go, and what goal they seek.

4. Mutual Context Network for Jointly Estimating Egocentric Gaze and Action
   Targets at addressing two coupled tasks of gaze prediction and action recognition in egocentric videos by exploring their mutual context: the information from gaze prediction facilitates action recognition and vice versa.

5. Ego-Exo: Transferring Visual Representations from Third-person to First-person Videos
   Introduce an approach for pre-training egocentric video models using large-scale third-person video datasets.

6. Ego-Topo: Environment Affordances from Egocentric Video
   Introduces a model for environment affordances that is learned directly from egocentric video. The main idea is to gain a human-centric model of a physical space (such as a kitchen) that captures (1) the primary spatial zones of interaction and (2) the likely activities they support.

7. You2Me: Inferring Body Pose in Egocentric Video via First and Second Person Interactions
   Proposes a learning-based approach to estimate the camera wearer's 3D body pose from egocentric video sequences.

8. Analysis of the hands in egocentric vision: A survey
   Reviews the literature that focuses on the hands using egocentric vision, categorizing the existing approaches into: localization (where are the hands or parts of them?); interpretation (what are the hands doing?); and application (e.g., systems that used egocentric hand cues for solving a specific problem)

9. The Roles of Vision and Eye Movements in the Control of Activities of Daily Living
   An early exploration and discussion on eye-hand coordination in human activities.

10. In the Eye of Beholder: Joint Learning of Gaze and Actions in First Person Video
    An egocentric daily activity dataset that provides annotations on action labels, gaze locations, hand masks for learning egocentric activities as well as human-object interactions. The most recent work on the EGTEA-GAZE line of research.

11. Home Action Genome: Cooperative Compositional Action Understanding
    A dataset that address activity/action recognition from multiple views with scene-graphs and detailed action annotation.

12. Ego-Exo: Transferring Visual Representations from Third-person to First-person Videos
    A first attempt on leveraging different views (egocentric vs. exocentric) for better activity/action understanding.

1.3 General-Purpose Vision

1.3.1 Topic Description

Design a vision system for general-purpose vision tasks. Propose benchmarks and metrics for evaluating the capability of generalization.

> How can we make vision systems that easily learn and adapt to new tasks and environments? Our goal is to reformulate visual learning as continual, opportunistically supervised, self-directed, and minimizing asymptotic error and sample complexity for unknown future tasks.
>
> — Derek Hoiem

1.3.2 Papers

1. Towards General Purpose Vision Systems
   Proposes a task-agnostic vision-language system that accepts an image and a natural language task description and outputs bounding boxes, confidences, and text.
   Video (3:40:00): Derek's Talk.

2. Perceiver: General Perception with Iterative Attention
   Builds a Transformers-based model that makes few architectural assumptions about the relationship between its inputs, but that also scales to hundreds of thousands of inputs.
   Video: Yannic's Explanation.

3. Zero-Shot Text-to-Image Generation
   DALL·E is a 12-billion parameter version of GPT-3 trained to generate images from text descriptions, using a dataset of text–image pairs.

4. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
   Introduces a new language representation model called BERT, which stands for Bidirectional Encoder Representations from Transformers.

5. Unifying Vision-and-Language Tasks via Text Generation
   Proposes a unified framework that learns different tasks in a single architecture with the same language modeling objective, i.e., multimodal conditional text generation.

6. Learning Transferable Visual Models From Natural Language Supervision
   OpenAI CLIP trains on 400 million images scraped from the web, along with text descriptions to learn a model that can connect the two modalities. The core idea is a contrastive objective combined with a large batch size. The resulting model can be turned into arbitrary zero-shot classifiers for new image and text tasks.
   Video: Yannic's Explanation.

7. How to represent part-whole hierarchies in a neural network
   Hinton proposes an imaginary system called GLOM, which has the potential to significantly improve the interpretability of the representations produced by transformer-like systems when applied to vision or language.
   Video: Yannic's Explanation.

8. HERO: Hierarchical Encoder for Video+Language Omni-representation Pre-training
   A general-purpose pretraining model for learning aligned video-language representation. Experiments cover video-text retrieval and videoQA.

9. Emerging Properties in Self-Supervised Vision Transformers
   A self-supervised transformer model for emerging properties, semantic segmentations.
   Video: Yannic's Explanation

1.4 Visual Reasoning

1.4.1 Topic Description

An intelligent agent should be able to solve problems beyond simply answering what is where. The ability to reason about the cause, the underlying mechanism, the future and use the results to guide our actions should be one of the defining features that make intelligent agents "intelligent".

> The study of vision must therefore include not only the study of how to extract from images the various aspects of the world that are useful to us, but also an inquiry into the nature of the internal reprensentations by which we capture this information and thus make it available as a basis for decisions about our thoughts and actions.
>
> — David Marr

1.4.2 Papers

- Tasks
  1. Measuring abstract reasoning in neural networks
     A Raven's Progressive Matrices dataset
  2. RAVEN: A Dataset for Relational and Analogical Visual rEasoNing
     Our work on Raven's Progressive Matrices
  3. ACRE: Abstract Causal REasoning Beyond Covariation
     Our work on causal Raven
  4. Machine Number Sense: A Dataset of Visual Arithmetic Problems for Abstract and Relational Reasoning
     An initial work on machine number sense
  5. PHYRE: A New Benchmark for Physical Reasoning
     Physical reasoning
  6. Rapid trial-and-error learning with simulation supports flexible tool use and physical reasoning
     Virtual tool game
- Methods
  1. Learning to Make Analogies by Contrasting Abstract Relational Structure
     A contrastive method in data preparation
  2. Learning Perceptual Inference by Contrasting
     Put the idea of contrast in modeling
  3. Abstract Spatial-Temporal Reasoning via Probabilistic Abduction and Execution
     Use probabilistic planning in solving the Raven problem
  4. Bridging machine learning and logical reasoning by abductive learning
     Abduction method to connect learning and reasoning
  5. Neural-Symbolic VQA: Disentangling Reasoning from Vision and Language Understanding
     Neuro-symbolic VQA with separate processes
  6. The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences From Natural Supervision
     Neuro-symbolic VQA with joint learning
  7. DeepProbLog: Neural Probabilistic Logic Programming
     Deep probabilistic logic
  8. DreamCoder: Growing generalizable, interpretable knowledge with wake-sleep Bayesian program learning
     Program synthesis synthesizer
  9. Closed Loop Neural-Symbolic Learning via Integrating Neural Perception, Grammar Parsing, and Symbolic Reasoning
     Another method of neuro-symbolic learning
  10. A HINT from Arithmetic: On Systematic Generalization of Perception, Syntax, and Semantics
      Add DreamCoder into the previous method

2 Cognition

2.1 Language and Thought

2.1.1 Topic Description

Languages vary tremendously in how they allow us to express ourselves. In some languages, you have to say when an event happened (past, present, future, etc.), while in others it is obligatory to say how you know about the event (you saw it, you heard about it), or what genders its participants were. In addition, languages just feel different from one another – some feel poetic while others feel brutal. Some things just don't sound right in certain languages, and some translations are harder than others to pull off. But are these differences meaningful? Do differences across languages cause substantive changes in the cognition of their speakers? We'll read some of the burgeoning research literature on these questions and consider how they can be answered with new empirical tools.

2.1.2 Introduction to Language and Thought

1. Story of Your Life
2. Boroditsky Ch0
3. Boroditsky Ch1
4. The Relation of Habitual Thought and Behavior to Language
5. Mentalese

2.1.3 Color

1. Russian blues reveal effects of language on color discrimination
2. Language, thought, and color: Whorf was half right
3. The Relationship Between Language and the Environment

2.1.4 Number

1. Boroditsky Ch5
2. Number as a cognitive technology: Evidence from Pirahã language and cognition
3. Verbal interference suppresses exact numerical representation
4. Representing Exact Number Visually Using Mental Abacus

2.1.5 Space

1. Boroditsky Ch3
2. Can language restructure cognition? The case for space
3. Spatial Reasoning in Tenejapan Mayans
4. Interaction between language and vision: It's momentary, abstract, and it develops
5. Remembrances of Times East

2.1.6 Social categories

1. Sex, Syntax, and Semantics
2. Grammatical Gender Effects on Cognition: Implications for Language Learning and Language Use
3. Thinking While Talking
4. Matched False-Belief Performance During Verbal and Nonverbal Interference

2.1.7 Events

1. Reconstruction of Automobile Destruction : An Example of the Interaction Between Language and Memory'
2. Subtle linguistic cues influence perceived blame and financial liability
3. Does language guide event perception? Evidence from eye movements

2.1.8 Words and Categories

1. Boroditsky Ch2
2. TOPIC... COMMENT
3. Languages Support Efficient Communication about the Environment: Words for Snow Revisited
4. Language can boost otherwise unseen objects into visual awareness
5. Language Is Not Just for Talking

2.1.9 Other forms of the hypothesis

1. What Is the Sapir-Whorf Hypothesis?
2. From "thought and language" to "thinking for speaking
3. The Whorfian Hypothesis: A Cognitive Psychology Perspective

Acknowledgements

This subsection of reading list is adapted from Stanford PSYCH 132.

2.2 Causal Cognition

2.2.1 Topic Description

Causality is central to our understanding of the world and of each other. We think causally when we predict what will happen in the future, infer what happened in the past, and interpret other people's actions and emotions. Causality is intimately linked to explanation – to answering questions about why something happened. In this discussion-based seminar class, we will first read foundational work in philosophy that introduces the main frameworks for thinking about causation. We will then read some work on formal and computational theories of causation that was inspired by these philosophical frameworks. Equipped with this background, we will study the psychology of causal learning, reasoning, and judgment. We will tackle questions such as: How can we learn about the causal structure of the world through observation and active intervention? What is the relationship between causal reasoning and mental simulation? Why do we select to talk about some causes over others when several causes led to an outcome? Toward the end of the course, we will discuss how what we have learned about causation in psychology may inform other fields of inquiry, such as legal science as well as machine learning and artificial intelligence.

2.2.2 Causality in Thought

1. Causality in Thought
   Keywords: *causal reasoning, causal judgment, causal decision making, causal attribution*.
2. Causal Conceptions in Social Explanation and Moral Evaluation: A Historical Tour
   Keywords: *causal judgment, attribution, moral judgment, social perception*.
3. Understanding "why": The role of causality in cognition
   A talk by Prof. Tobias Gerstenberg.

2.2.3 Philosophy of Causation

1. Ideas about causation in philosophy and psychology
2. Causation: Interactions between Philosophical Theories and Psychological Research
3. Two Concepts of Causation
4. Contrastive causation.
5. Causal–explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions
6. Counterfactual Theories of Causation
7. Counterfactuals

2.2.4 Theories of Causation

1. Representing causation
2. Causes and Explanations: A Structural-Model Approach—Part I: Causes
3. Causal Models
4. For want of a nail: How absences cause events
5. Graded Causation and Defaults
6. What Is Wrong with Bayes Nets?

2.2.5 Causal Perception

1. Causation From Perception
2. Perceptual causality and animacy
3. Visual Adaptation of the Perception of Causality
4. Causation without realism
5. Causal perception and causal cognition
6. Time reordered: Causal perception guides the interpretation of temporal order
7. The perception of causality in infancy

2.2.6 Causal Learning

1. A Theory of Causal Learning in Children: Causal Maps and Bayes Nets
2. Are Causal Structure and Intervention Judgments Inextricably Linked? A Developmental Study
3. Theory-based causal induction
4. Beyond covariation
5. Structure and strength in causal induction
6. Intuitive experimentation in the physical world

2.2.7 Causal Reasoning

1. Programs as Causal Models: Speculations on Mental Programs and Mental Representation
2. The Problem of Causal Selection
3. Cause and Norm
4. Normality and actual causal strength
5. Seeing Versus Doing: Two Modes of Accessing Causal Knowledge
6. The problem of variable choice
7. Sensitive and Insensitive Causation
8. The role of causality in judgment under uncertainty
9. Norm theory: Comparing reality to its alternatives

2.2.8 Causal Judgement

1. The causal asymmetry
2. Indicators of causal agency in physical interactions: The role of the prior context
3. Reconciling intuitive physics and Newtonian mechanics for colliding objects
4. Eye-Tracking Causality
5. Concepts in a Probabilistic Language of Thought

2.2.9 Causality in AI and in the Law

1. Building machines that learn and think like people
2. Explanation in Artificial Intelligence: Insights from the Social Sciences
3. The seven tools of causal inference, with reflections on machine learning
4. Common-Sense Causation in the Law
5. Contrastive causation in the law
6. Causation in Legal and Moral Reasoning
7. Causation and Responsibility

Acknowledgements

This subsection of reading list is adapted from Stanford PSYCH 291.

2.3 Communication, Intentionality, and the Origins of Language

2.3.1 Topic Description

How did language evolve to become a ubiquitous, definitional part of human life? What relationship does children's early language have to their understanding of intentionality and other methods of non-verbal communication? This seminar will survey theoretical and experimental work on the foundations of human language, communication, and intentionality, with the goal of understanding what we know and what questions are still open. Areas of focus include developmental work on communication; whether early language use is referential/intentional and whether early words are general or particular; and research on language evolution and animal communication.

2.3.2 Joint attention and pedagogy

1. The capacity for joint visual attention in the infant
2. Gaze Following in Human Infants Depends on Communicative Signals
3. Communication-induced memory biases in preverbal infants
4. Natural pedagogy as evolutionary adaptation

2.3.3 Is early language understanding intentional?

1. Infants' Contribution to the Achievement of Joint Reference
2. One-Year-Old Infants Appreciate the Referential Nature of Deictic Gestures and Words
3. Understanding the abstract role of speech in communication at 12 months
4. Twelve-month-old infants recognize that speech can communicate unobservable intentions

2.3.4 Early language: Associative or referential?

1. At 6–9 months, human infants know the meanings of many common nouns
2. Early word-learning entails reference, not merely associations
   (a) Response to Sloutsky: Taking development seriously: Theories cannot emerge from associations alone
   (b) Theories about 'theories': where is the explanation? Comment on Waxman and Gelman
3. Do Both Pictures and Words Function as Symbols for 18- and 24-Month-Old Children?
4. Thinking of Things Unseen: Infants' Use of Language to Update Mental Representations

2.3.5 Ape theory of mind and precursors of language

1. Does the chimpanzee have a theory of mind?
2. Does the Chimpanzee have a Theory of Mind? 30 years later
3. Differential Sensitivity to Human Communication in Dogs, Wolves, and Human Infants
4. Ape gestures and language evolution

2.3.6 Social cooperation and language

1. Origins of Human Communication
2. Altruistic Helping in Human Infants and Young Chimpanzees
3. Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis
4. Collaboration encourages equal sharing in children but not in chimpanzees

2.3.7 The recursion debate

1. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?
2. What's special about the human language faculty?
3. Recursive syntactic pattern learning by songbirds
4. Songbirds possess the spontaneous ability to discriminate syntactic rules

2.3.8 Language evolution

1. The Origin of Speech
2. Language evolution: consensus and controversies
3. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language
4. Language Evolution by Iterated Learning With Bayesian Agents

2.3.9 Language as shaped by cognition

1. A Parsing Theory of Word Order Universals
2. Language as shaped by the brain
3. Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change
4. The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning

2.3.10 Language as shaped by communication

1. Coordinating perceptually grounded categories through language: A case study for colour
2. Word lengths are optimized for efficient communication
3. The communicative function of ambiguity in language
4. Predicting Pragmatic Reasoning in Language Games

2.3.11 Case studies: Color, lightness, and kinship

1. Color naming reflects optimal partitions of color space
2. Modeling the emergence of universality in color naming patterns
3. The Relationship Between Language and the Environment: Information Theory Shows Why We Have Only Three Lightness Terms
4. Kinship Categories Across Languages Reflect General Communicative Principles

Acknowledgements

This subsection of reading list is adapted from Stanford PSYCH 293 and UC Berkeley PSYCH 290

2.4 Interacting “like a human being”

2.4.1 Topic Description

In recent years the psychological processes underlying cooperation and communication have received considerable attention both from a developmental and a comparative, evolutionary perspective. Experimental studies have shown that several group-living primate species are able to coordinate their actions flexibly in cooperative problem-solving tasks by either carrying out identical or complementary actions to achieve their objectives. Similarly, recent research on the evolutionary roots of language has highlighted both the complexity of primate communication and its difference from human communication. Simultaneously, scholars investigating human sociality from a sociological, anthropological and linguistic perspective have began to uncover different layers of order in terms of structure, timing, sequential and pragmatically consequential unfolding of social interaction in adult human beings. In this seminar we will read papers from these very diverse lines of research to better understand the key building blocks of human sociality and we will discuss how we can systematically investigate what it means to interact “like a human being”.

2.4.2 Introduction

1. The Neglected Situation
2. The Origin of Speech

2.4.3 Cooperative principle, Relevance Theory and ontogeny of speech acts

1. Précis of Relevance: Communication and Cognition
2. The ontogenesis of speech acts
3. Origins of human communication

2.4.4 Common Ground, Shared Cooperative Activities and Shared Intentionality

1. Grounding in communication
2. Shared Cooperative Activity
3. Understanding and sharing intentions: The origins of cultural cognition
4. On the Human "Interaction Engine"

2.4.5 Earlier attempts to study social interaction

1. Interaction process analysis; a method for the study of small groups
2. A classification of illocutionary acts
3. The repertoire of nonverbal behavior: Categories, origins, usage, and coding
4. Interaction ritual: Essays in face to face behavior

2.4.6 Spatial arrangements and F-formation

1. Studies in Personal Space
2. Spatial Organization in Social Encounters : the F-formation System
3. Proxemics [and Comments and Replies]
4. What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy

2.4.7 Turn-Taking, Sequence Organization, and Storytelling

1. Sequence organization in interaction: Volume 1: A primer in conversation analysis
2. Why is conversation so easy?
3. A simplest systematics for the organization of turn-taking for conversation
4. Storytelling in conversation

2.4.8 Referring, Word Selection, Entitlement, Epistemics

1. Conceptual pacts and lexical choice in conversation
2. Some practices for referring to persons in talk-in-interaction: A partial sketch of a systematics
3. Contingency and Action: A Comparison of Two Forms of Requesting
4. Epistemics in Action: Action Formation and Territories of Knowledge

2.4.9 Repair

1. Monitoring and self-repair in speech
2. The preference for self-correction in the organization of repair in conversation
3. Repair After Next Turn: The Last Structurally Provided Defense of Intersubjectivity in Conversation
4. Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn-Beginning

2.4.10 Cooperation and Prosociality and Social Manipulation

1. How is human cooperation different?
2. Social norms and human cooperation
3. Social manipulation in nonhuman primates: Cognitive and motivational determinants
4. Darwin, deception, and facial expression

Acknowledgements

This subsection of reading list is adapted from UC San Diego COGS 260.

3 Language

3.1 Language Models

3.1.1 Topic Description

The language model performs the probability distribution over sentences. According to Goldberg, Yoav (2017), "Language modeling is the task of assigning a probability to sentences in a language. [...] Besides assigning a probability to each sequence of words, the language models also assigns a probability for the likelihood of a given word (or a sequence of words) to follow a sequence of words". Language modeling is essential to many important natural language processing tasks in real-world applications such as machine-translation and automatic speech recognition. For these reasons, language modeling plays a crucial role general AI research.

3.1.2 General Language Modeling

1. A Neural Probabilistic Language Model. In 2003 *Journal of Machine Learning Research*.
2. Recurrent neural network based language model. In *INTERSPEECH 2010*.
3. Statistical Language Models Based On Neural Networks. *TOMÁŠ MIKOLOV's PhD Thesis*.
4. Sequence to Sequence Learning with Neural Networks. In *NIPS 2014*.
5. Pointer Networks. In *NIPS 2015*.
6. Attention Is All You Need. In *NIPS 2017*.

3.2 Grammar Induction

3.2.1 Topic Description

Parsing aims to uncover the intrinsic structure (e.g., a constituent or dependency structure) of a sentence or an image. Such syntactic structures have been found useful in downstream tasks such as semantic parsing, relation extraction, and machine translation. Unsupervised parsing aims to learn a parser from sentences or images that have no annotation of their correct parse trees. Despite its difficulty, unsupervised parsing is an interesting research direction because of its capability of utilizing almost unlimited unannotated data.

3.2.2 Language-Vision Joint Parsing

1. Joint Image-Text News Topic Detection and Tracking by Multimodal Topic And-Or Graph
2. Joint Video and Text Parsing for Understanding Events and Answering Queries
3. Visually Grounded Compound PCFGs
4. Video-aided Unsupervised Grammar Induction

3.2.3 Language Parsing

1. Unsupervised Natural Language Parsing (Introductory Tutorial)
2. Unsupervised Neural Dependency Parsing
3. Corpus based induction of syntactic structure: Models of dependency and constituency
4. Enhancing unsupervised generative dependency parser with contextual information
5. Compound Probabilistic Context-Free Grammars for Grammar Induction
6. Unsupervised recurrent neural network grammars

3.3 Natural Language Understanding

3.3.1 Topic Description

"There is a growing need for systems that can understand and generate natural language in applications that require substantial amounts of knowledge as well as reasoning capabilities. Most current implemented systems for natural language understanding (NLU) are decoupled from any reasoning processes, which makes them narrow and brittle. Furthermore, they do not appear to be scalable in the sense that the techniques used in such systems do not appear to generalize to more complex applications. While significant work has been done in developing theoretical underpinnings of systems that use knowledge and reasoning (*e.g.*, development of models of linguistic interpretation using abductive reasoning, intention recognition, formal models of dialogue, formal models of lexical and utterance meaning, and utterance planning), it has often proved difficult to utilize such theories in robust working systems.
Another major barrier has been the vast amount of linguistic and world knowledge needed. However, there is now significant progress in compiling the required knowledge, using manual and increasingly automated techniques for ontology and grammar learning. But even as these resources become available, we still lack some key conceptual and computational frameworks that will form the foundation for effective scalable natural language systems, e.g., in terms of incremental processing, dialogical alignment or pragmatics. The collection of researchers who face the challenges involved in scaling human language technology is growing in conjuction with greater efforts to develop systems that robustly interact with users in intuitive and conversational ways."

3.3.2 Language Grounding

1. Learning Hierarchical Space Tiling for Scene Modeling, Parsing and Attribute Tagging
2. Experience Grounds Language
3. Recent Advances in Natural Language Inference: A Survey of Benchmarks, Resources, and Approaches

3.4 Knowledge Base, Knowledge Graphs, and Commonsense

3.4.1 Topic Description

"More and more knowledge graphs are constructed for private use, *e.g.*, Siri, Alexa, or public use, *e.g.* DBpedia, Wikidata. While techniques to automatically construct KGs from existing Web objects exist (*e.g.*, scraping Web tables), there is still room for improvement. Initially, constructing KGs from existing datasets was considered an engineering task with some ad-hoc approaches, however, scientific methods with more declarative-oriented techniques have recently emerged. In particular, several mapping languages for describing rules to construct knowledge graphs and processors to execute those rules emerged. Addressing the challenges related to KG construction requires both the investigation of theoretical concepts and the development of tools and methods for their evaluation."
"Commonsense knowledge graphs (CSKGs) are sources of background knowledge that are expected to contribute to downstream tasks like question answering, robot manipulation, and planning. The knowledge covered in CSKGs varies greatly, spanning procedural, conceptual, and syntactic knowledge, among others. CSKGs come in a wider variety of forms compared to traditional knowledge graphs, ranging from (semi-)structured knowledge graphs, such as ConceptNet, ATOMIC, and FrameNet, to the recent idea to use language models as knowledge graphs. As a consequence, traditional methods of integration and usage of knowledge graphs might need to be expanded when dealing with CSKGs. Understanding how to best integrate and represent CSKGs, leverage them on a downstream task, and tailor their knowledge to the particularities of the task, are open challenges today. The workshop on CSKGs addresses these challenges, by focusing on the creation of commonsense knowledge graphs and their usage on downstream commonsense reasoning tasks."

3.4.2 Knowledge Graph

1. Domain-specific knowledge graphs: A survey
2. Exploiting Linked Data and Knowledge Graphs in Large Organisations
3. DBpedia: A Nucleus for a Web of Open Data
4. Learning Entity and Relation Embeddings for Knowledge Graph Completion
5. Random walk inference and learning in a large scale knowledge base
6. Improving Learning and Inference in a Large Knowledge-base using Latent Syntactic Cues

3.4.3 Commonsense Knowledge

1. A theory of commonsense knowledge
2. ConceptNet—a practical commonsense reasoning tool-kit
3. Socialiqa: Commonsense reasoning about social interactions

3.5 Natural Language Generation

3.5.1 Topic Description

In general, while Natural Language Understanding takes up the understanding of the texts based on grammar, semantic, and intent, Natural Language Generation produces human language, in either spoken or written form. They are subsections of the natural language processing domain.

3.5.2 Text Generation

1. I2T: Image Parsing to Text Description
2. Auto-Encoding Variational Bayes
3. A Hierarchical Neural Autoencoder for Paragraphs and Documents
4. Generating Sentences from a Continuous Space

3.5.3 Automatic Evaluation

1. Towards holistic and automatic evaluation of open-domain dialogue generation

3.6 Cognitive Theory in Language

3.6.1 Topic Description

Natural Language is closely linked with cognitive theories in terms of reasoning (*e.g.*commonsense reasoning and pragmatics reasoning), theory-of-mind (ToM) modeling, emergence of language, *etc.* This seminar will go through early literature on linguistics, cognitive science, and neural science, exploring how human learn, think, and then communicate.

3.6.2 Pragmatics and Discourse

1. Meaning in Language: An Introduction to Semantics and Pragmatics Oxford Handbook in Linguistics
2. Pragmatics and Discourse
3. Implicature Stanford Encyclopedia of Philosophy.

3.6.3 Theory-of-Mind

1. Word meaning in minds and machines
2. Revisiting the Evaluation of Theory of Mind through Question Answering. In *EMNLP 2019*.

3.6.4 Emergence of Language

1. Emergence of language Nature Physics
2. Iterated learning: A framework for the emergence of language
3. The emergence of language from embodiment
4. Emergence of language with multi-agent games: Learning to communicate with sequences of symbols
5. Multi-agent cooperation and the emergence of (natural) language

References

- International Workshop On Knowledge Graph Construction
- Workshop on Common Sense Knowledge Graphs
- ACL Workshop on Scalable Natural Language Understanding

4 Multi-agent Systems

4.1 Multi-agent Reinforcement Learning

1. Decentralised multi-agent reinforcement learning for dynamic and uncertain environments
   A decentralised MARL algorithm enhanced by a prediction mechanism that provides accurate information regarding up-coming changes in the environment.

2. Markov games as a framework for multi-agent reinforcement learning
   This paper explores the Markov game formalism as a mathematical framework for reasoning about multi-agent environments.

3. Multi-agent reinforcement Learning in sequential social dilemmas
   A study of cooperation and competition in sequential social dilemmas.

4. Reinforcement social learning of coordination in networked cooperative multiagent systems
   The interactions among agents are described by two representative topologies: the small-world and the scale-free network.

5. Markov security games: learning in spatial security problems
   This paper presents a preliminary investigation of modelling spatial aspects of security games within the context of Markov games.

6. Emergence of grounded compositional language in multi-agent populations
   This paper investigates if, and how, grounded compositional language can emerge as a means to achieve goals in multi-agent populations.

7. Transfer learning in multi-agent reinforcement learning domains
   This work proposes a novel method for transfer learning (the process of reusing knowledge from past tasks in order to speed up the learning procedure in new tasks.) in multi-agent reinforcement learning domains.

8. Continuous adaptation via meta-Learning in nonstationary and competitive environments
   This paper casts the problem of continuous adaptation into the learning-to-learn framework.

9. Multi-agent inverse reinforcement learning
   This paper introduces the problem of multiagent inverse reinforcement learning, where reward functions of multiple agents are learned by observing their uncoordinated behavior.

4.2 Game Theory and Nash Equilibria

1. Games, Strategies, and Decision Making (Joseph E. Harrington, Jr.)
   This book introduces core concepts with a minimum of mathematics in order to give you insights into human behavior.

2. Game networks
   Game networks, a novel representation for multi-agent decision problems.

3. Computing equilibria for two-person games *A self-contained survey of algorithms for computing Nash equilibria of two-person games given in normal form or extensive form*

4. Chapter 2 Computation of equilibria in finite games
   An overview of the methods for numerical computation of Nash equilibria.

5. Cheap talk, coordination, and entry
   We show how costless, nonbinding, nonverifiable communication (cheap talk) can achieve partial coordination among potential entrants into a natural-monopoly industry.

4.3 Agent-based Modelling

1. Agent-Based Models
   This short book explains what agent-based modeling is. It also warns of some dangers and describes typical ways of doing agent-based modeling. Finally, it offers a range of examples from many of the social sciences.

2. Collective Action
   This article examines collective action, focusing on the role of social interactions, conflict, and the dynamics of interpersonal influence in shaping collective identities and interests.

3. A physical analogue of the Schelling model

4. The strength of social interactions and obesity among women

5. Experimental study of critical-mass fluctuations in an evolving sandpile
   Demonstrates that real, finite-size sandpiles may be described by models of self-organized criticality.

6. Social force model for pedestrian dynamics
   Suggests that pedestrian motion can be described by a simple social force model for individual pedestrian behavior.

5 Literature Review

5.1 Document Preparation: LaTeX

One needs to be proficient in preparing documents in LaTeX. Some LaTeX style requirements could be found in the "LaTeX Author Guidelines for CVPR/ICCV Proceedings", *e.g.*, http://cvpr2021.thecvf.com/sites/default/files/2020-09/cvpr2021AuthorKit_2.zip

1. Online LaTeX Editor: Overleaf
    - LaTeX documentation, quickstart/guide and Overleaf guides: https://www.overleaf.com/learn
    - LaTeX templates: https://www.overleaf.com/latex/templates
    - Bibtex: https://www.overleaf.com/learn/latex/Bibliography_management_with_bibtex
2. Local LaTeX editors:
    - VS Code with LaTeX Workshop extension: Installation Guide (Almost real-time compilation with 'Auto Save' turned on)
    - A live typesetter (mainly for MacOS): Texpad
3. Email etiquette: Basics (slide), Format, Tips, and Structure (w/ video)
4. Grammar check: Grammarly

5.2 Literature Review

It is important and necessary to write a literature review before the start of a research project. Without comprehensive understanding in the related research areas, one could not draw inspirations from previous work and bring novel ideas. When completing a literature review, you will be confident to answer the question: "What is your motivation to do this project?"

One may found that papers in the field of computer vision often start with Section 1: *Introduction* and Section 2: *Related Works*. A brief literature review could be a combination of these two sections, without introduction to a newly proposed approach though. One could further discuss some potential future work instead.

In the *Introduction* section, the authors tell the big story of their work (Ref: Derek Hoiem, Jennifer Widom):

1. Why is the problem important?
2. Why is the problem hard?
3. Why hasn't it been solved before?
4. What is your main contribution?
5. What is your approach?
6. What is significant/worthy about your contribution and approach?
7. How does your work fit into the broader picture?

The final paragraph or subsection of the *Introduction* section is usually "Summary of Contributions," which lists the major contributions in bullet form.

In the *Related Works* section, audience expect to learn how the new method in questions differs from an already published method (ref: BEEHIVE). The authors describe previous work in the related research areas and place the new method's contributions to the field in this context.

How to Read a Paper: "Paper reading skills are put to the test in doing a literature survey."

It is common that you may still have questions on some concepts and approaches in the existing literature, especially learning about an unfamiliar area. Feel free to write them down in an extra question section and ask your supervisor/senior students.

5.3 Tips for Finding Related Works

1. Given a paper, read its cited papers to expand your paper library.
2. Given a paper, read the papers citing it (find them using Google Scholar) to expand your paper library.
3. For the first authors in paper library, check their homepage, Google Scholar page, and lab website to find their recent following works.

5.4 Sample Surveys

There are surveys in various areas of artificial intelligence. They are comprehensive but could be very long. Take it easy and skim to learn how to structure a literature review.

1. Dark, Beyond Deep: A Paradigm Shift to Cognitive AI with Humanlike Common Sense
2. Transformers in Vision: A Survey
3. Deep Learning for 3D Point Clouds: A Survey
4. Visual Affordance and Function Understanding: A Survey