The Visual Narrative Engine
A Computational Model of the Visual Narrative Parallel Architecture

Chris Martens, North Carolina State University
Rogelio E. Cardona-Rivera, University of Utah rogelio@cs.utah.edu
Neil Cohn, Tilburg University http://rogel.io

@recardona
How do people understand visual narrative?
We don’t know the process that underlies visual story understanding.
We don’t know the process that underlies visual story understanding.
We don’t know the process that underlies visual story understanding.

- AI is broadly functionalist
We don’t know the process that underlies visual story understanding.

• AI is broadly functionalist …they are rational hypotheses, and that’s okay
We don’t know the process that underlies visual story understanding.

- AI is broadly functionalist …they are rational hypotheses, and that’s okay.
We don’t know the process that underlies visual story understanding.

• AI is broadly functionalist...they are rational hypotheses, and that’s okay
  human unconstrained

• Knowledge-lean Story Understanding: Document Analysis

• Knowledge-rich Story Generation: Narrative-theoretic Heuristic Search Planning

• Neurosymbolic Understanding+Generation: Benchmarked Commonsense Reasoning
We don’t know the process that underlies visual story understanding.

• AI is broadly functionalist
  ▪ Remarkable progress on story understanding

Knowledge-rich Story Understanding

By the 1990s, we knew...

Wilensky et al. — people predict which goals and subsequent plans explain observed actions by characters

Norvig — scripts are important for generating knowledge-based inferences

Mueller — spatiotemporal reasoning constrains story inferences

Lehnert — means-ends (causal) and hierarchical (purposive) reasoning contribute the most to a person’s memory of a story

Black and Bower — hierarchical problem solving is key for inferencing and understanding

Winston — it is possible to combine these in a principled manner
Knowledge-rich Story Understanding

By the 1990s, we knew…

Wilensky—et al.—Norvig—Mueller—Lehnert—Black and Bower—Winston— it is possible to combine these in a principled manner

The Psychology Survey says…

They were all right about the concepts!

But not right about the procedures.
Knowledge-rich Story Understanding

By the 1990s, we knew...

Wilensky et al.—Norvig—Mueller—Lehnert—Black and Bower—Winston

people predict which goals and subsequent plans explain observed actions by characters

scripts are important for generating knowledge-based inferences

spatiotemporal reasoning constrains story inferences

means-ends (causal) and hierarchical (purposive) reasoning contribute the most to a person's memory of a story

hierarchical problem solving is key for inferencing and understanding

it is possible to combine these in a principled manner

The Psychology Survey says...

They were all right about the concepts!

But not right about the procedures.

Fun fact:

Neurosymbolic approaches are discovering the same

• Events
• Goals
• Characters
• Scripts
Knowledge-rich Story Understanding

By the 1990s, we knew…

Wilensky et al.—Norvig—Mueller—Lehnert—Black and Bower—Winston—people predict which goals and subsequent plans explain observed actions by characters. Scripts are important for generating knowledge-based inferences. Spatiotemporal reasoning constrains story inferences. Means-ends (causal) and hierarchical (purposive) reasoning contribute the most to a person’s memory of a story. Hierarchical problem solving is key for inferencing and understanding. It is possible to combine these in a principled manner.

The Psychology Survey says…

They were all right about the concepts! But not right about the procedures.

Fun fact:

Neurosymbolic approaches are discovering the same:

- Events
- Goals
- Characters
- Scripts

Another fun fact:

The 80s knew what we had to do

Although the HST theory suggests the form of the products of comprehension in the reader’s memory, it is seriously deficient in not spelling out the moment-by-moment process by which the reader arrives at those representational products. This deficit is the primary focus for the theoretical work in the future. Although we have no process theory at present, we will indicate some of the considerations and issues that must be resolved in arriving at a process model for story comprehension.
We don’t know the process that underlies visual story understanding.

- AI is broadly functionalist
  - Remarkable progress on story understanding

We don’t know the process that underlies visual story understanding.

- AI is broadly functionalist
  - Remarkable progress on story understanding
- Story psychology has offered process-level accounts
We don’t know the process that underlies visual story understanding.

- AI is broadly functionalist
  - Remarkable progress on story understanding
- Story psychology has offered process-level accounts
  - Too abstract

Rogelio E. Cardona-Rivera and R. Michael Young; Desiderata for a Computational Model of Human Online Narrative Sensemaking. In the Working Notes of the 2019 AAAI Spring Symposium on Story-enabled Intelligence, Stanford, CA, USA, 2019.
We don’t know the process that underlies visual story understanding.

• AI is broadly functionalist
  ▶ Remarkable progress on story understanding

• Story psychology has offered process-level accounts
  ▶ Too abstract

• Soapbox: Must bring these together
  ▶ Biologically-plausible structural models

This paper is an existence proof: we can describe the (visual) story understanding procedures mechanically.
This paper is an existence proof: we can describe the (visual) story understanding procedures mechanically. We may need to rely on simulation.
We discretize a narration into its constituent event structure.

And what happens next is anybody’s guess.
Parallel Interfacing Narrative Semantics
Dual (Syntax/Semantics) Process Reasoning

Graphic Structure

Narrative Categories ↔ Narrative Constituents ↔ Structural Revision

Semantic Memory ↔ Semantic Expectancies ↔ Situation Model

Access → Prediction → Updating
Visual Narrative Grammar

Graphic Structure

Narrative Structure

Initial₁ Prolongationₐ Peak₅ Release₆

Pr₂ Pr₃ Pr₄

Event Structure

Event

Preparation Head Coda

IMPACT(D, G)₅ BE(D, ON(G))₆

Spatial/Referential Structure

Air

Dragon

Ground

LAUNCH(D, FROM(A)₁, VIA(A)₂,₃,₄, TO(G)₅)
The Visual Narrative Engine
Model of the combined VNG+PINS = Visual Narrative Parallel Arch.

• Research Question: Can we describe procedures to match posited interfaces?
The Visual Narrative Engine
Model of the combined VNG+PINS = Visual Narrative Parallel Arch.

• Research Question: Can we describe procedures to match posited interfaces?
Scene Graphs
A Representation from Computer Vision
The Visual Narrative Engine
Model of the combined VNG+PINS = Visual Narrative Parallel Arch.

- Research Question: Can we describe procedures to match posited interfaces?
Hierarchical Task Networks
A Representation from Automated Planning
Representations are Compatible!

Binary Literals
- wearing(?girl, ?dress)
- holding(?girl, ?balloon)

Unary Literals
- flowery(?dress)
- standing(?girl)
- red(?balloon)
- pink(?balloon)

Rogelio E. Cardona-Rivera and Boyang Li; *PlotShot: Generating Discourse-constrained Stories around Photos*. In Proceedings of the 12th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-16), pages 2-8, Burlingame, CA, USA, 2016.
Open question: What’s the bridge?

- Research Question: Can we describe procedures to match posited interfaces?
Open question: What’s the bridge?

Hypothesis: Both?

Hypothesis: Hierarchical Plan Recognition?

Hypothesis: Hierarchical Planning?
We don’t know the process that underlies visual story understanding.
We don’t know the process that underlies visual story understanding.

- AI is broadly functionalist
  - Remarkable progress on story understanding
- Story psychology has offered process-level accounts
  - Too abstract
- Soapbox: Must bring these together
  - Biologically-plausible structural models

rogelio@cs.utah.edu
http://rogel.io
@recardona