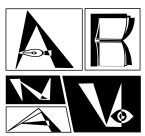
A COMPUTATIONAL MODEL FOR VISUAL NARRATIVE COMPREHENSION

Yi-Chun (Rimi) Chen

ychen74@ncsu.edu

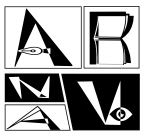
Department of Computer Science
North Carolina State University

26 August 2021



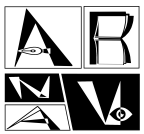
Agenda

1. Background
   - Introduction
   - Questions
   - Comic
   - Literature
2. Findings
   - Features
   - Transitions
   - Graphical Style
   - Generation
3. Project Design
   - Model
   - Stages
   - Timeline
   - Q & A



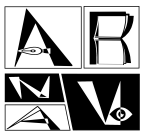
Introduction

- Visual Media transport ideas with efficiency, effectiveness, influences.



Introduction

- Visual Media transport ideas with efficiency, effectiveness, influences.
- When narratives are complex, often sequences of images: films, games, comics, etc.



Introduction

- Visual Media transport ideas with efficiency, effectiveness, influences.
- When narratives are complex, often sequences of images: films, games, comics, etc.
- Comics are a representative form of sequential visual narratives, easier to be created, widely published through various platforms.

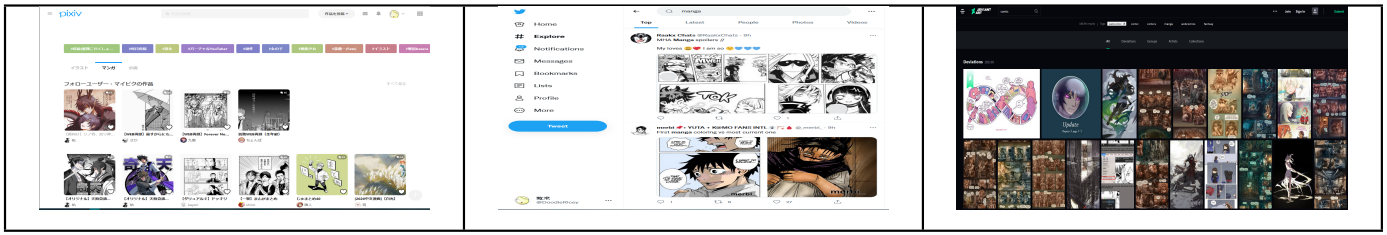


Table: Comics in social platforms.

Introduction

- Visual Media transport ideas with efficiency, effectiveness, influences.
- When narratives are complex, often sequences of images: films, games, comics, etc.
- Comics are a representative form of sequential visual narratives, easier to be created, widely published through various platforms.

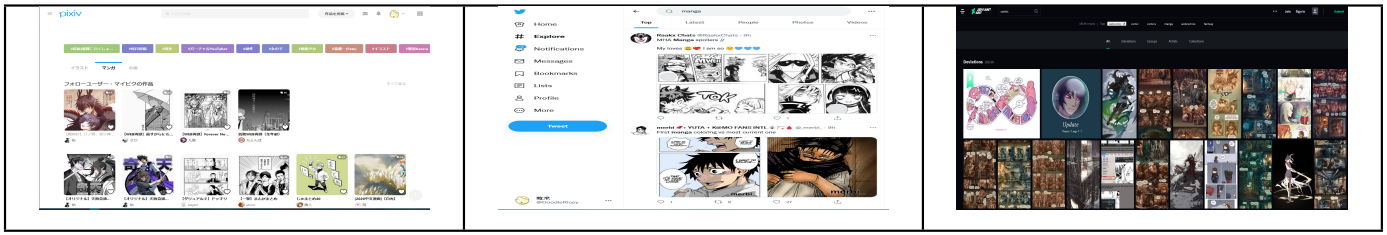


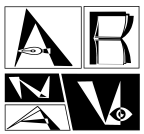
Table: Comics in social platforms.

- How to automatically understand them?

A visual narrative comprehension model!

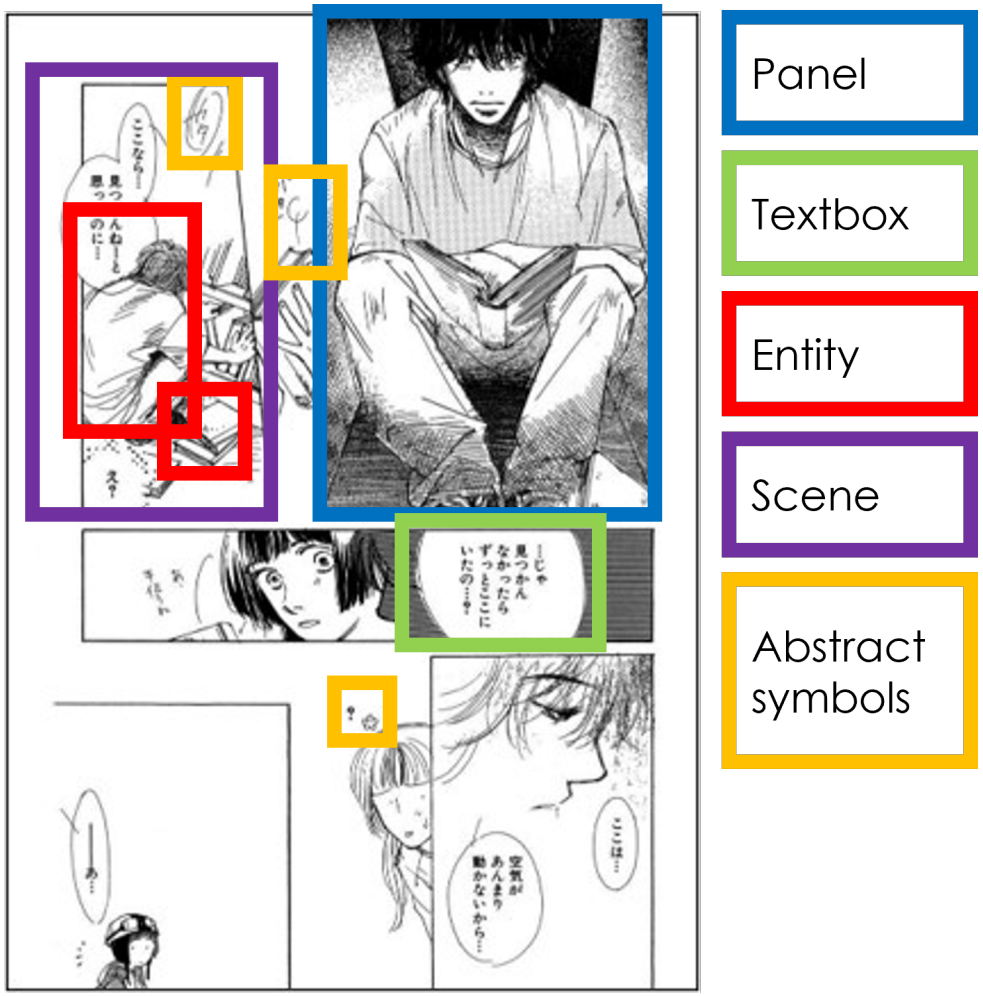
Research Questions

1. What influences comic understanding?
2. What are the components of a comic comprehension model?
3. How to integrate various comics features that benefit comprehension?
4. How to evaluate the comprehension model?



Comic

Figure: Elements in comics



Comic Sequence

Table: Sample sequence in western comics and Japanese manga

Literature Review

Comic Theories

- Panel transitions [1]
- Visual Narrative Grammar (VNG) [2]

Conceptual Models

- Scene Perception & Event Comprehension Theory (SPECT) [3]
- Parallel Interfacing Narrative-Semantics (PINS) [4]

Computational Models

- Visual Narrative Engine (VNE) [5]
- Hierarchical LSTM [6]

Dataset

- COMICS [6]
- Manga109 [7]

Panel transitions

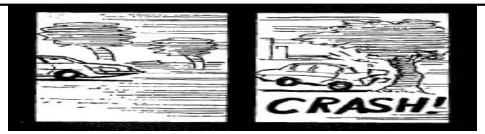
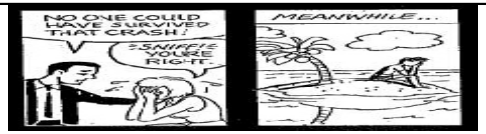
Moment	Action	Subject
		
Scene	**Aspects**	**Non**
		

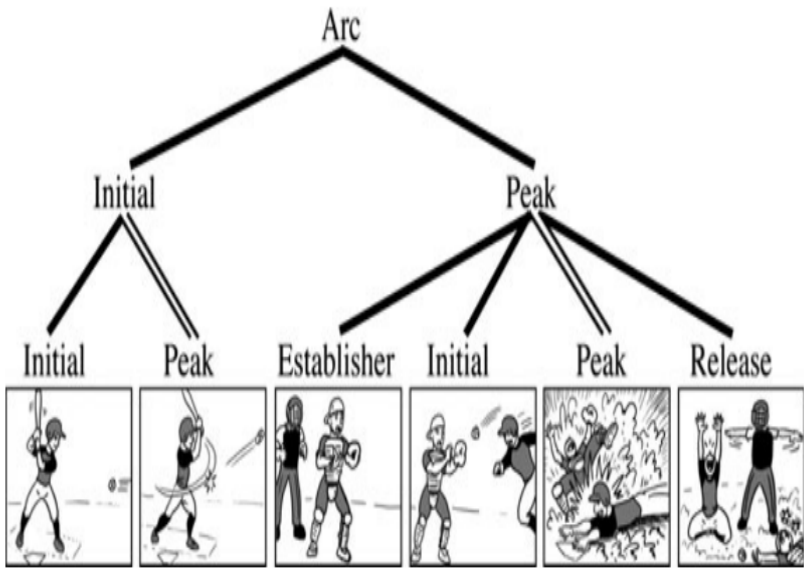
Table: Transitions

Features	Advantages
■ Transitions 1 Moment-to-Moment 2 Action-to-Action 3 Subject-To-Subject 4 Scene-To-Scene 5 Aspects-To-Aspects 6 Non-sequitur	■ Captured operations on reader's attention ■ Modeled visual representations

Table: Transitions info

Visual Narrative Grammar

Figure: Narrative Structure

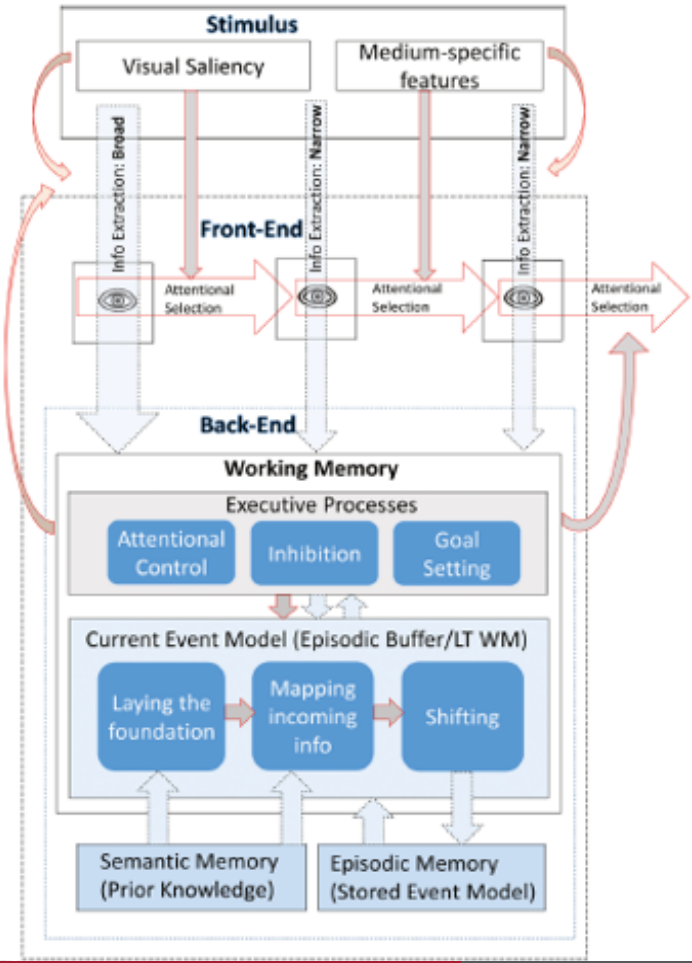


Features
■ Categories 1 Establisher (E) 2 Initial (I) 3 Prolongation (L) 4 Peak (P) 5 Release (R) ■ Phase (E)-(I(L))-P-(R)
Advantages
■ Structuralized visual narrative

Table: VNG info

Scene Perception & Event Comprehension Theory (SPECT)

Figure: SPECT



Features

- Visual Stimulus
- Front-end (attention selection)
- Back-end (semantic processing)

Advantages

- Decomposed visual narrative comprehension
- Formulated cognitive process

Visual Narrative Engine

Figure: Scene graphs of VNE

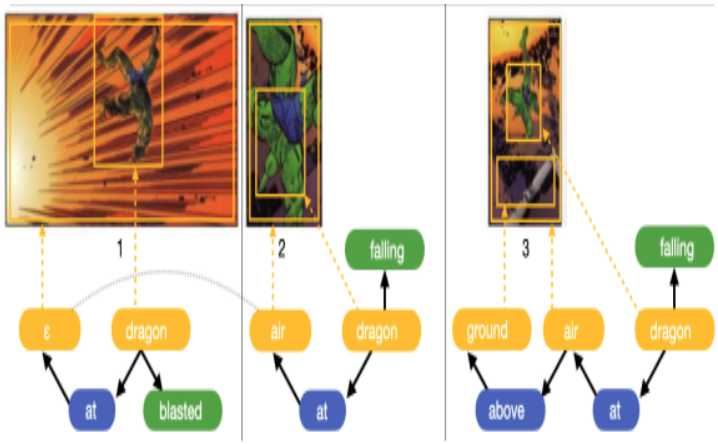
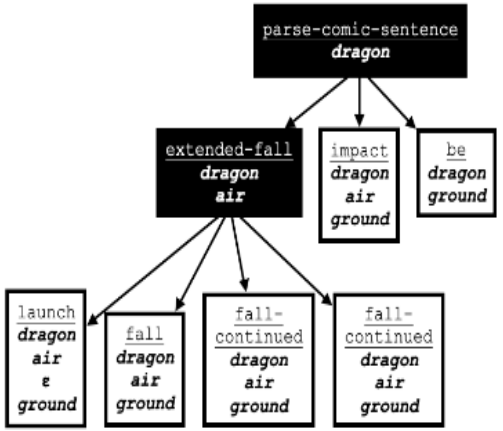


Figure: A hierarchical plan

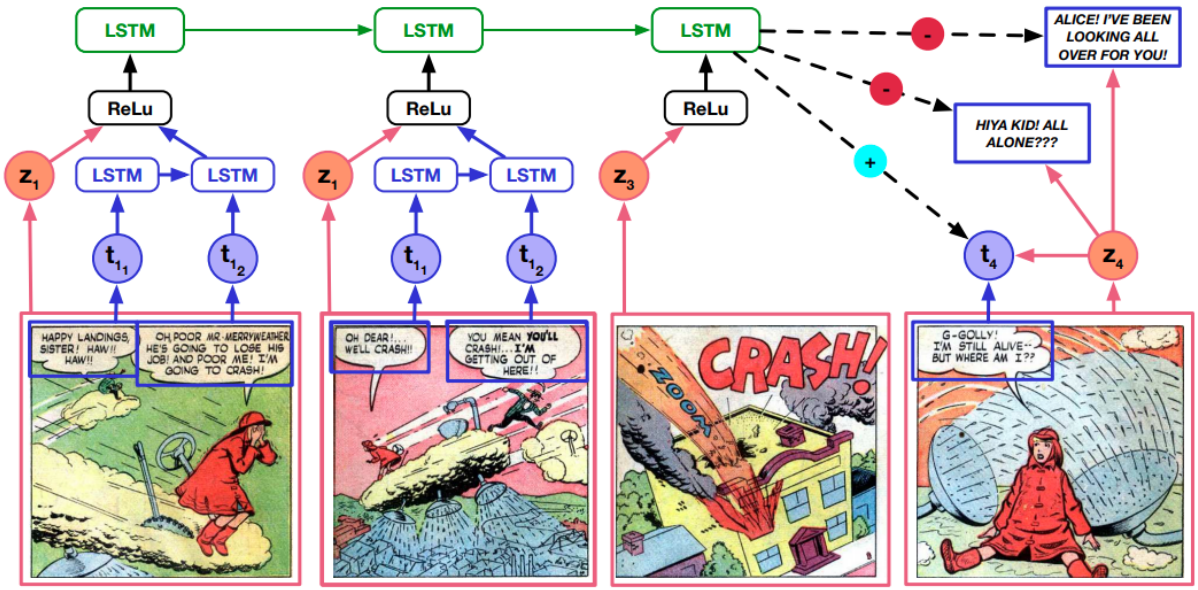


Features	Advantages
■ Event model ■ Spatial structures	■ Modeled semantics ■ Structuralized content representations

Table: VNE info

Hierarchical LSTM

Figure: Hierarchical LSTM



Features	Advantages
■ Multi-modalities of comics ■ Closure tasks	■ Framed as vision task ■ Dealt with massive data

Table: Hierarchical LSTM info

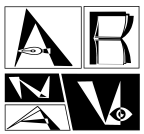
Datasets

COMICS	Manga109
3,948 American comics – 1.2 million panels ■ Publicly available ■ 1938 – 1954 ■ diverse genres	109 titles of Japanese manga 21,142 pages – ■ Commercially published ■ 1970s - 2010s ■ 12 genres

Table: Comics resources

Comic Features

- Motivation
  Investigate the possible features that influence comic comprehension
- Correspond research Question
  What features of comics influence understanding?



Comic Features: Method and Results

Figure: Manga layout

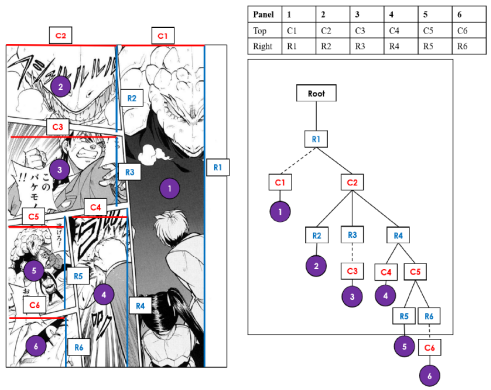


Figure: Cloze-test results

Tasks	Base easy	Base hard	W/ style easy	W/ style hard
Text-only,text	42.4	51.5	36.0	36.6
Image-text,text	42.4	45.4	34.7	37.2
Image-only,visual	51.4	47.9	**70.3**	**67.8**
Image-text,visual	47.6	44.7	**60.9**	**55.8**

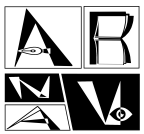
Features	Contributions
■ Layout complexity ■ Reading orders ■ Text annotations ■ Image information formula	■ Enhanced dataset ■ Combined perceptual features ■ Measured information image carries

Table: Frist findings: Features influences understanding.

Inter-panel Transitions & Content

- Motivation
  Study the relevance between panel transitions and the story content

- Correspond research Questions
  - What are the components of a comic comprehension model?
  - Whether the discourse related to story content?



Inter-panel Transitions & Content: Method and Results

Figure: Layered CNN

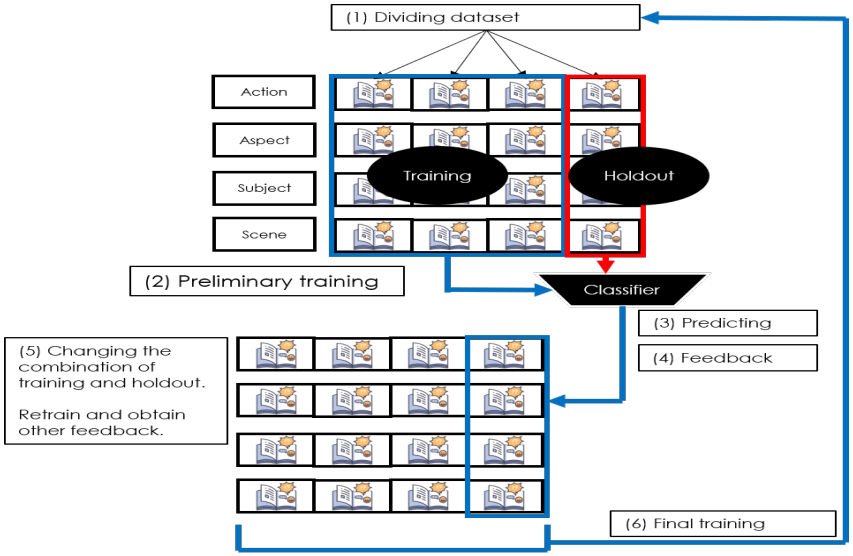


Figure: Inter-rater reliability

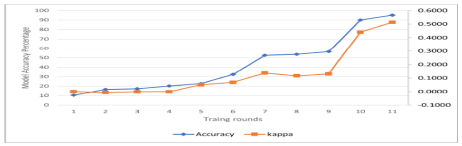
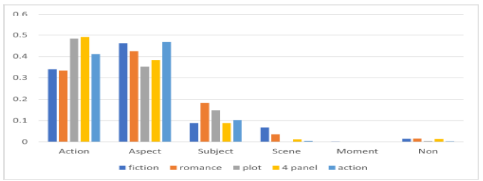


Figure: Transition distributions



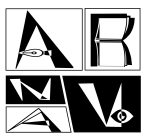
Features	Contributions
■ Transition labeling ■ Genre analysis	■ Analyzed genres and transitions ■ Automated labeling

Table: Second findings: Inter-panel relations affect content In

Intra-panel Relations & Graphical Style

- Motivation
  Investigated the relation between intra-panel content and drawing style.

- Correspond research Questions
  - What features influence comic understanding?
  - What are the elements of a comic comprehension model?
  - If drawing style changed will semantics still be preserved if the scene relation is kept?



Intra-panel Relations & Graphical Style: Method and Results

Figure: Style transfer framework

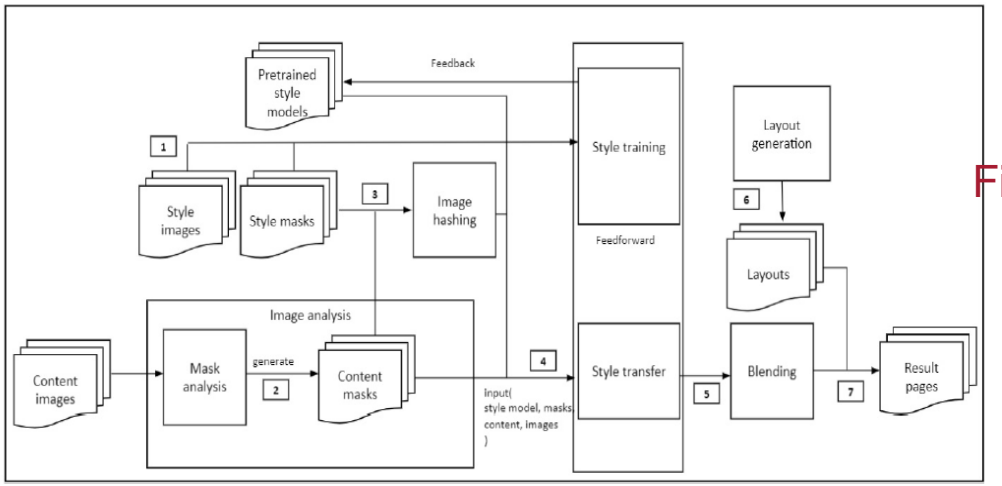
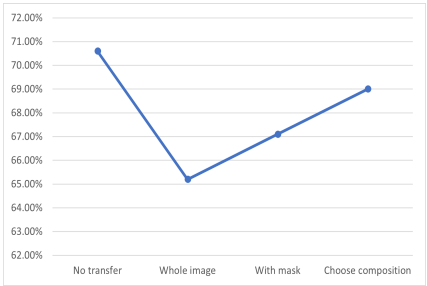


Figure: Semantic preservation

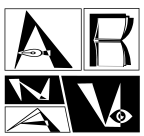


Features	Contributions
■ Style transfer with/without perceptual masks ■ Image composition analysis	■ Demonstrated content semantic preservation ■ Set transfer framework with masks

Table: Third findings: Study the Visual representations and Intra-panel Rela-

Comic Generation

- Motivation
  - Experiment comic theories
  - Reduce cost of finding cases
  - Integrate analyzed components of comics
- Correspond research Questions
  - How comic theories model comic content?
  - Does generated comic form understandable sequences?
  - How comic elements interact?



Comic Generation: Method and Results

Figure: Comic generator

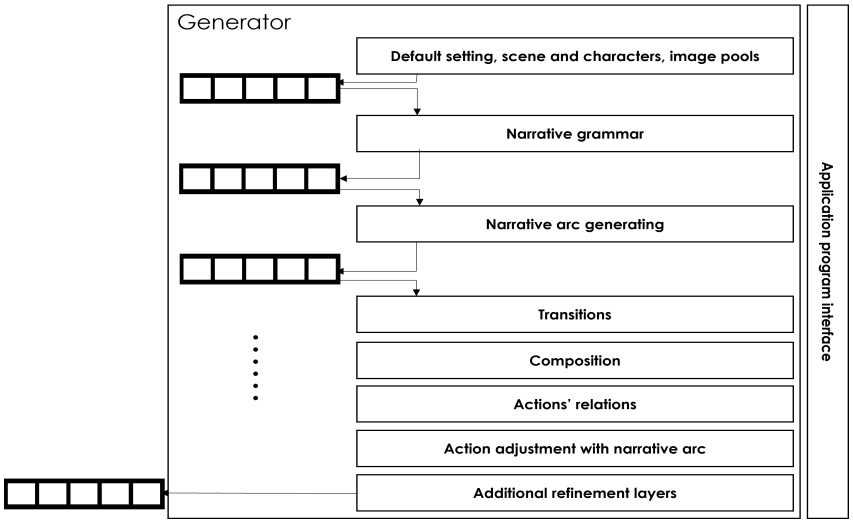


Figure: Generated comic #1

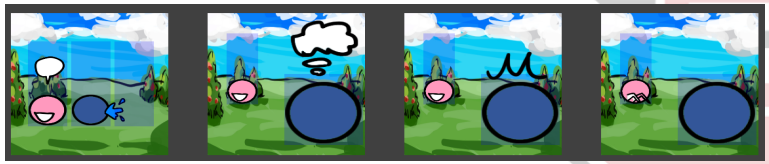
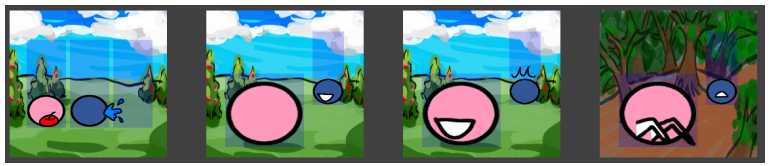


Figure: Generated comic #2

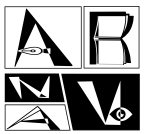


Features	Contributions
■ Adjustable generator ■ Comic theories' application ■ Abstract comic metaphors	■ Provided model and API ■ Provided a graphical set for comic generation ■ Implementedtheories

Table: Side Project: Applying Comic Theories to Generate Comics

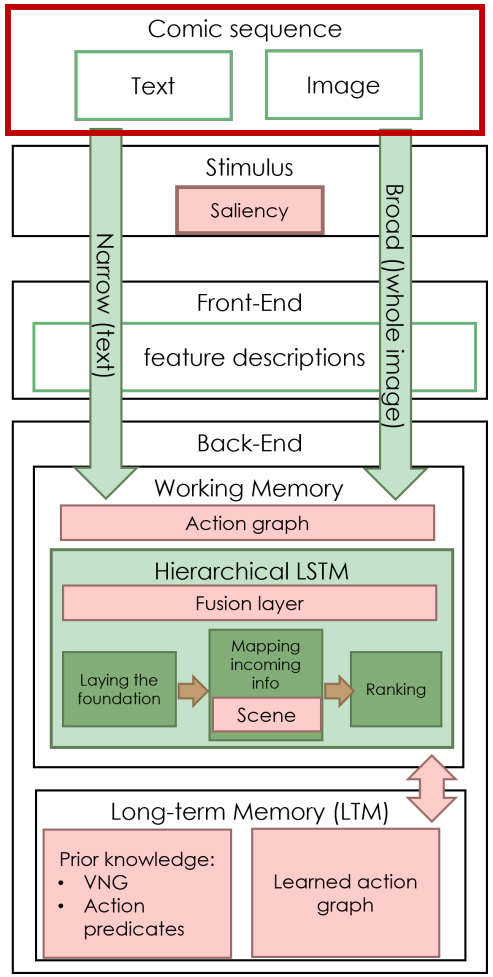
Design Motivation

- Combine advantages to improve comprehension model
  - Develop content awareness through scene representation
  - Encode multi-modal features into metrics to fit massive data processing
  - Employ formalized cognitive model



Cognitive Model

Figure: Modified SPECT



Prior Knowledge: External Resources

Figure: Manga109 annotations

- **face** : Face of a character
  - Character ID
- **body** : Body of a character
  - Character ID
- **text** : Typed text and some handwritten text
  - text content
- **frame** : Frame
  - No additional information

- Scene
  - Scene ID
- Action
  - Action ID
  - Character ID

Figure: VerbNet predicates

FRAMES REF KEY

NP V NP

EXAMPLE "Doug removed the smudges."

SYNTAX Agent V Theme

SEMANTICS cause(Agent, E) location(start(E), Theme, ?Source) not(location(end(E), Theme, ?Source))

NP V NP PP.source

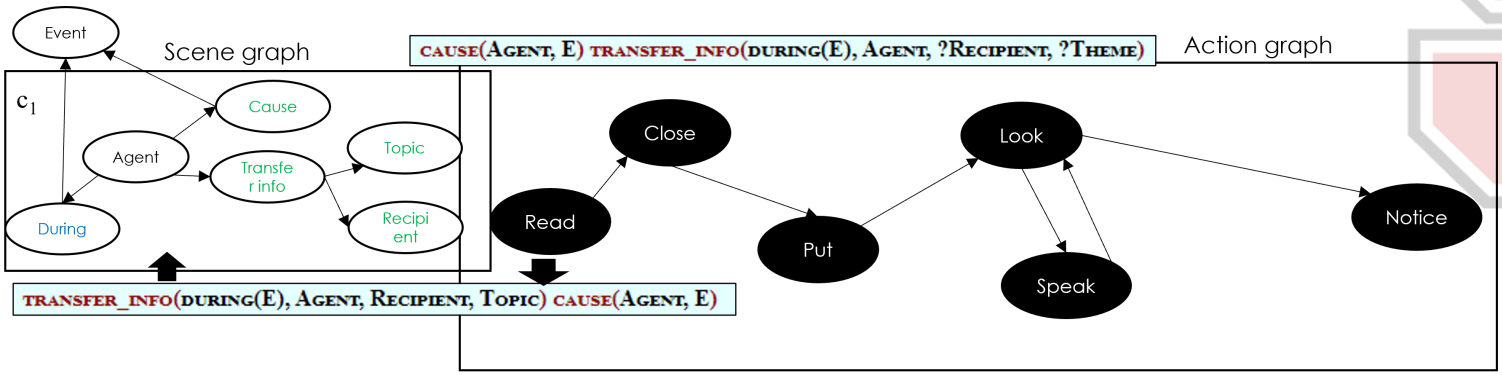
EXAMPLE "Doug removed the smudges from the tabletop."

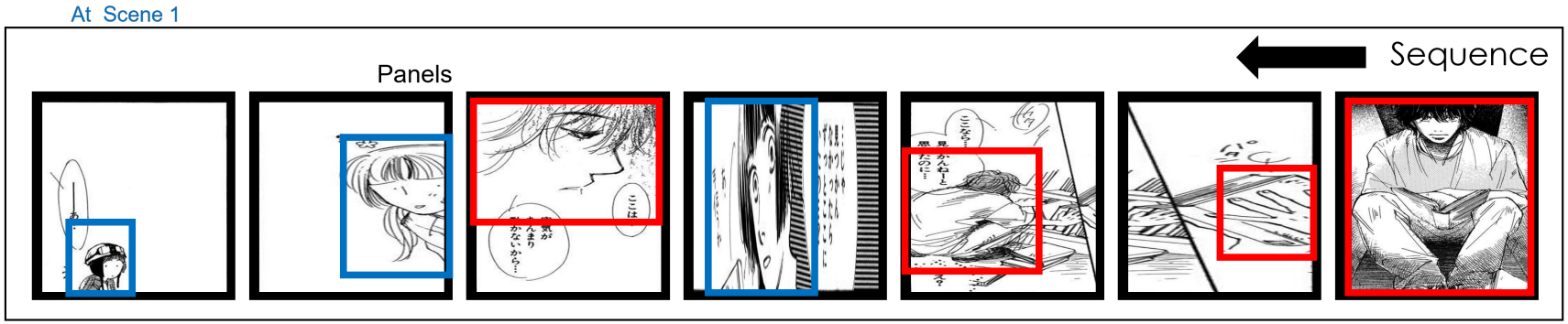
SYNTAX Agent V Theme {{+src}} Source

SEMANTICS cause(Agent, E) location(start(E), Theme, Source) not(location(end(E), Theme, Source))

Action graphs& Scene representation

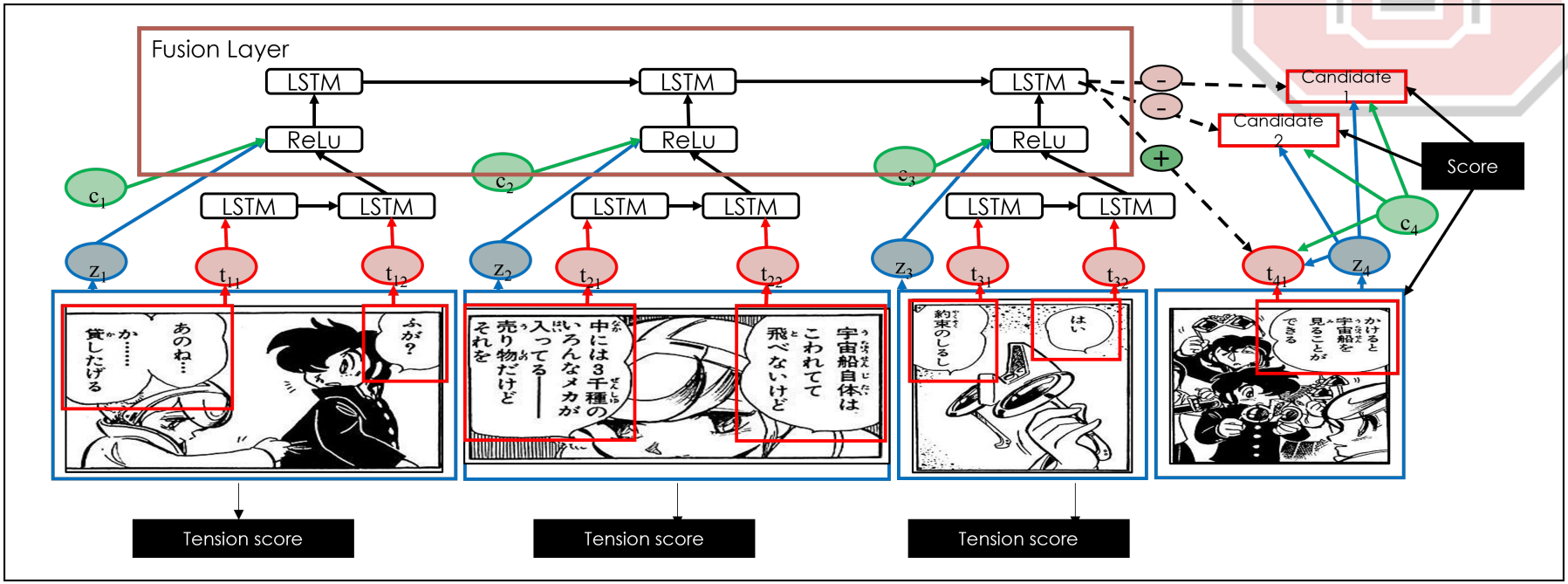
Figure: Action graph and scene representation





Hierarchical LSTM

Figure: Proposed hierarchical LSTM



Possible Tasks & Challenges

Action Graphs and Scene Representations

- Detect actions
  - Action detection in images
- Action graphs
  - Graph structure implementation
- Scene graphs
  - Processing VerbNet predicates

Narrative Structure

- Rank narrative tension of comic panels
  - Ranking algorithm
  - Lower level image features
  - Mapping between Structure and Narrative

Integration

- Integrate scenes with LSTM
  - Graph to one-dimensional representation
- Integrate with SPECT
  - Evaluation

Stages

Stage 1: Prototype

- Implement model structures
- Demonstrate the hypothesis
- Test with known generated examples
- Expected result:
  1. Cloze-test on generated examples
  2. Compare performance between with/without scene knowledge

Stage 2: Minimum Deliverable

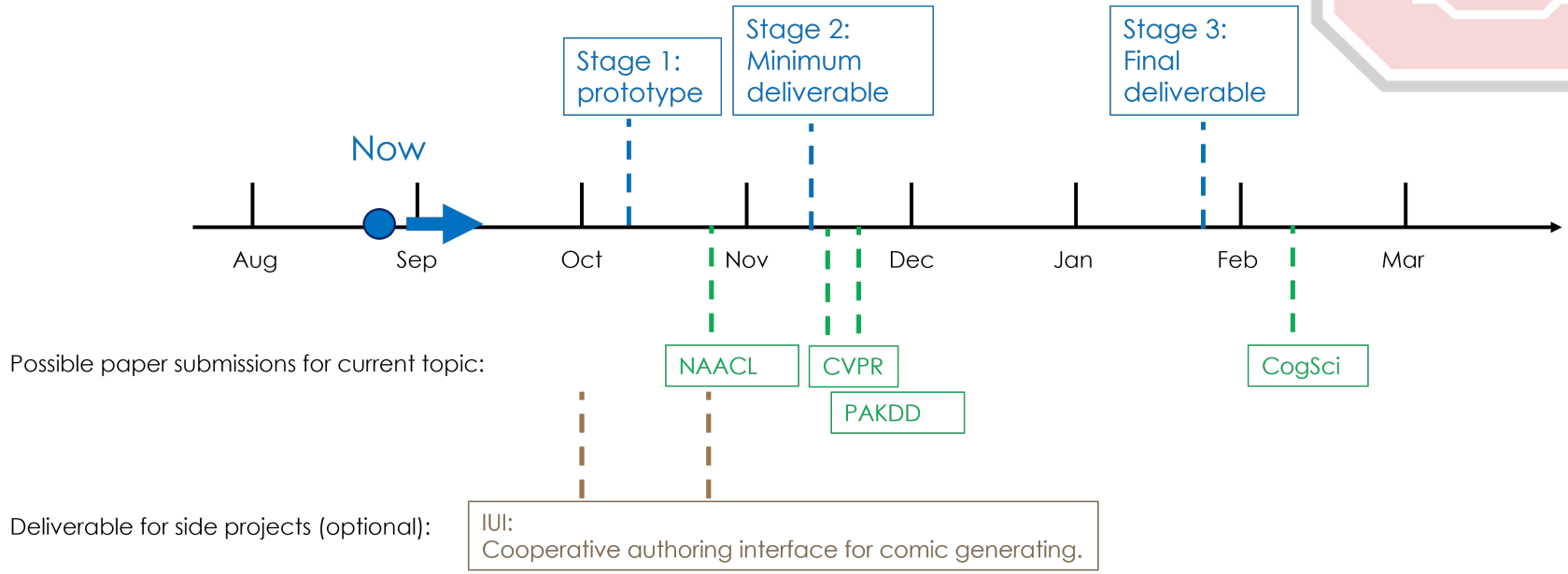
- Annotate two or three genres (each 2 or 3 books) of real comics
- Adapt model with annotated data
- Expected result:
  1. Compare performance between new model and original ones

Stage 3: Final Result

- Action Detection for comics' images

Timeline

Figure: Timeline



References I

- Scott McCloud and Mark Martin. **Understanding comics: The invisible art**. Vol. 106. Kitchen sink press Northampton, MA, 1993.
- Neil Cohn. “Visual narrative structure”. In: **Cognitive science** 37.3 (2013), pp. 413–452.
- Lester C Loschky et al. “The scene perception & event comprehension theory (SPECT) applied to visual narratives”. In: **Topics in cognitive science** 12.1 (2020), pp. 311–351.
- N Cohn. “Your Brain on Comics: A Cognitive Model of Visual Narrative Comprehension.”. In: **Topics in cognitive science** (2019).
- Chris Martens et al. “The Visual Narrative Engine: A Computational Model of the Visual Narrative Parallel Architecture”. In: **8th Annual Conference on Advances in Cognitive Systems**. 2020.

References II


Mohit Iyyer et al. “The amazing mysteries of the gutter: Drawing inferences between panels in comic book narratives”. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. 2017, pp. 7186–7195.


Yusuke Matsui et al. “Sketch-based manga retrieval using manga109 dataset”. In: **Multimedia Tools and Applications** 76.20 (2017), pp. 21811–21838.

Thank you!

Q & A