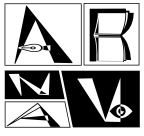
A COMPUTATIONAL MODEL FOR VISUAL NARRATIVE COMPREHENSION

Yi-Chun (Rimi) Chen

ychen74@ncsu.edu

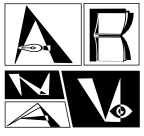
Department of Computer Science
North Carolina State University

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

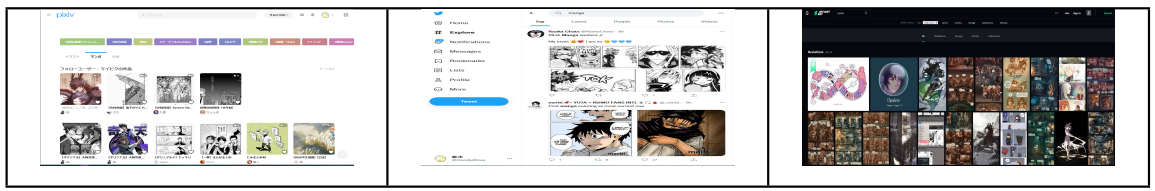
- Visual narratives are powerful tools.
  - Telling complex tales with compact timeline(Brehmer et al., 2016).
  - Engagement and entertainment experiences (Lankow et al., 2012).
  - Learn complicated idea quicker (3M_Meeting_Network, 1998).

Introduction

- Visual narratives are powerful tools.
  - Telling complex tales with compact timeline(Brehmer et al., 2016).
  - Engagement and entertainment experiences (Lankow et al., 2012).
  - Learn complicated idea quicker (3M_Meeting_Network, 1998).
- Visual narratives have many forms such as films, games, comics, etc.

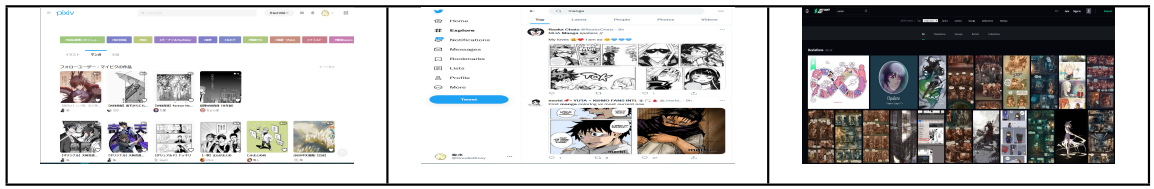
Introduction

- Visual narratives are powerful tools.
  - Telling complex tales with compact timeline(Brehmer et al., 2016).
  - Engagement and entertainment experiences (Lankow et al., 2012).
  - Learn complicated idea quicker (3M_Meeting_Network, 1998).
- Visual narratives have many forms such as films, games, comics, etc.
- Comic is one of the representative forms.
  - Express features of narrative such as time, tension, etc.
  - Less complex than video, not interactive like games.
  - Multi-modality on the understanding.
  - Popular culture (Lopes, 2006).



Introduction

- Visual narratives are powerful tools.
  - Telling complex tales with compact timeline(Brehmer et al., 2016).
  - Engagement and entertainment experiences (Lankow et al., 2012).
  - Learn complicated idea quicker (3M_Meeting_Network, 1998).
- Visual narratives have many forms such as films, games, comics, etc.
- Comic is one of the representative forms.
  - Express features of narrative such as time, tension, etc.
  - Less complex than video, not interactive like games.
  - Multi-modality on the understanding.
  - Popular culture (Lopes, 2006).



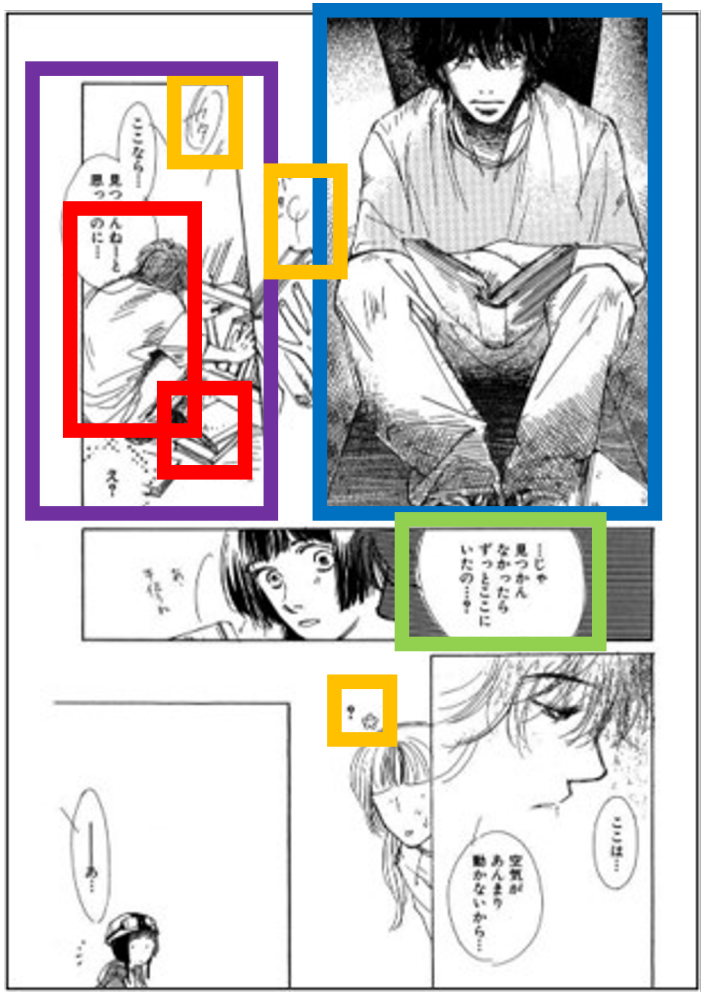
- How to automatically understand their content?

A computational model that comprehends visual narratives.

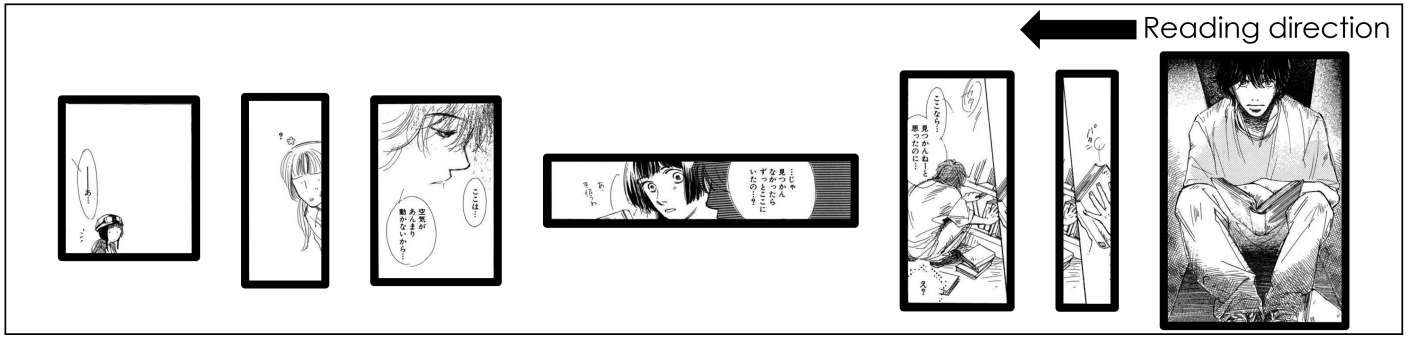
Research Questions

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

Elements in Comics



Comic Sequence



Literature Review

Vocabulary of comics

- Panel transitions (McCloud & Martin, 1993)
- Visual Narrative Grammar (VNG) (Cohn, 2013)

Cognitive Models

- Scene Perception & Event Comprehension Theory (SPECT) (Loschky et al., 2020)
- Parallel Interfacing Narrative-Semantics (PINS) (Cohn, 2019)

Computational Models

- Visual Narrative Engine (VNE) (Martens et al., 2020)
- Hierarchical LSTM (Iyyer et al., 2017)

Dataset

- COMICS (Iyyer et al., 2017)
- Manga109 (Matsui et al., 2017)

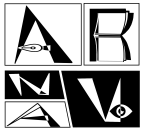
Datasets

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

Comic Features

- Motivation
  Investigate the possible features that influence comic comprehension

- Corresponding research questions
  What features of comics influence understanding?

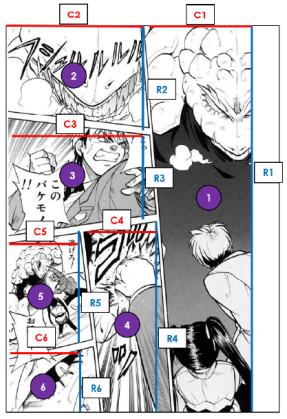


Comic Features: Method and Results

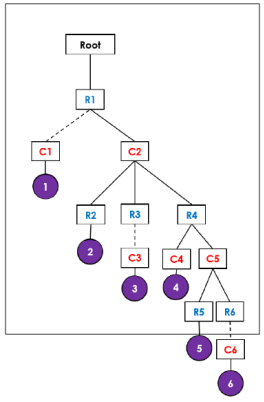
Features	Contributions
■ Layout structure and complexity ■ Reading orders ■ Text annotations ■ Image information score formula	■ Enhanced dataset ■ Combined perceptual features ■ Introduced a baseline for study manga comprehension

Comic Features: Layout Structure and Complexity

Manga layout



Panel	1	2	3	4	5	6
Top	C1	C2	C3	C4	C5	C6
Right	R1	R2	R3	R4	R5	R6



Irregular layout

Manga examples	Non-regular layout	Simple layout	Average layout	Complex layout	Complex score
AosugiruHaru	0.26	0.06	0.33	0.33	0.8
BakuretsuKungFuGirl	0.42	0.11	0.21	0.26	1.0
BEMADER_P	0.04	0.13	0.41	0.45	0.4

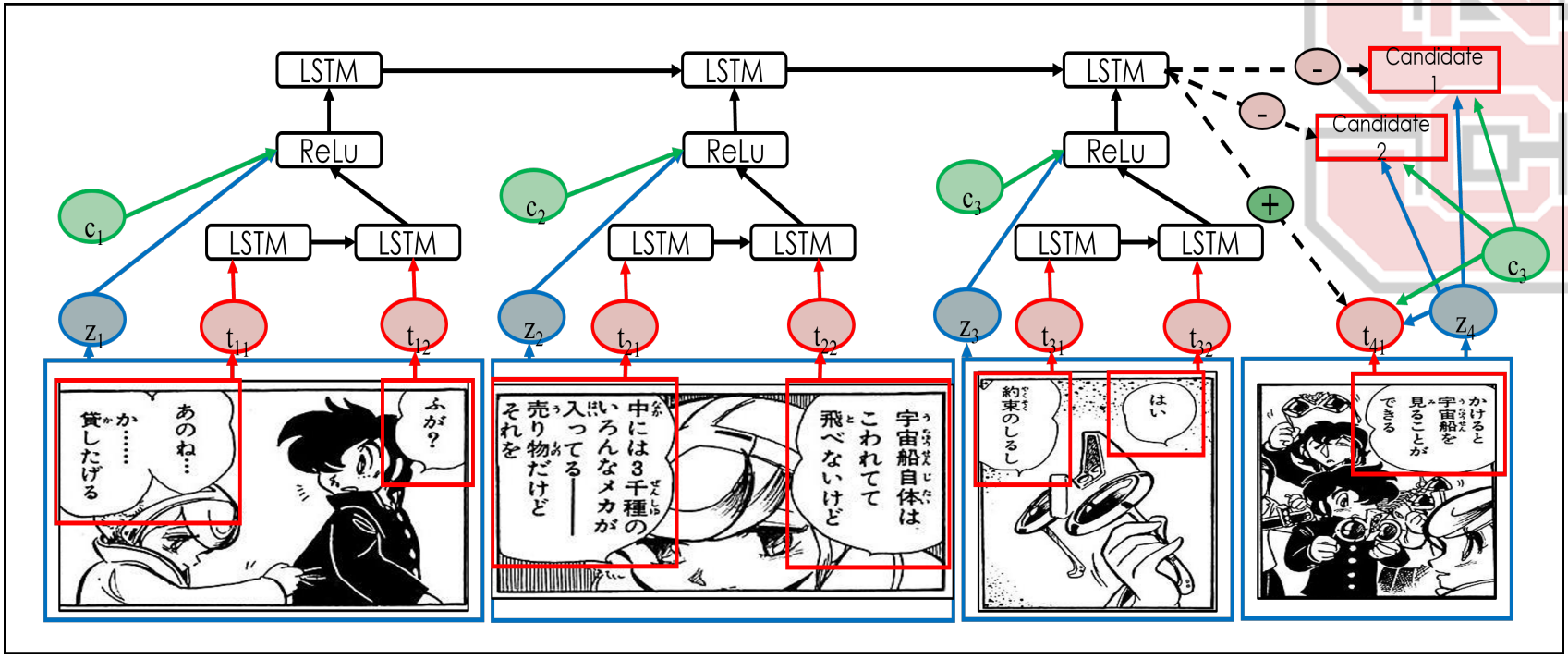
Comic Features: Image Expression Formula

Image information score formula:

$$information\_score = w_1 \times IA + w_2 \times LS + w_3 \times PN + w_4 \times TN$$

- $IA = \frac{image\ area\ in\ panel}{sum\ of\ image\ areas}$
- LS = layout complexity score
- PN = panel density, $\frac{panels\ in\ a\ page}{average\ used\ number\ of\ panels}$
- TN = textbox density, textbox ratio in pages
- $\{w_1, w_2, w_3, w_4\}$ = weights

Comic Features: Model



- Text features combined with image features in the hierarchical LSTM to get context representation.
- Concatenate perceptual features to add new information of context.

Comic Features: Evaluation and Results

Cloze-test:

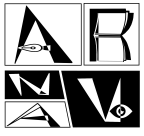
- Given sequence context, predict correct endings out of *N* candidates, 3 in our settings.
- Types of closure tasks:
  1. Predict visual content in ending panel.
  2. Predict text content in ending panel.
- Feature settings: text only, visual only, image & text
- Difficulties of test sets:
  1. Easy: Candidates from further panel sets (wider page number ranges)
  2. Hard: Candidates from closer panel sets (narrow page number ranges)

Tasks	Base easy	Base hard	W/ Info score easy	W/ Info score 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**

Inter-panel Transitions & Content

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

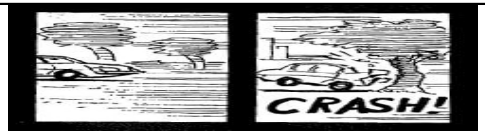
- Corresponding research questions
  - What are the components of a comic comprehension model?
  - Is story content influence the composition features?



Inter-panel Transitions & Content: Method and Results

Features	Contributions
■ Panel transition labeling ■ Panel transition uses of genres	■ Provided annotations for comic narrative studies. ■ Automated labeling.

Inter-panel Transitions & Content: Panel Transitions (McCloud & Martin, 1993)

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

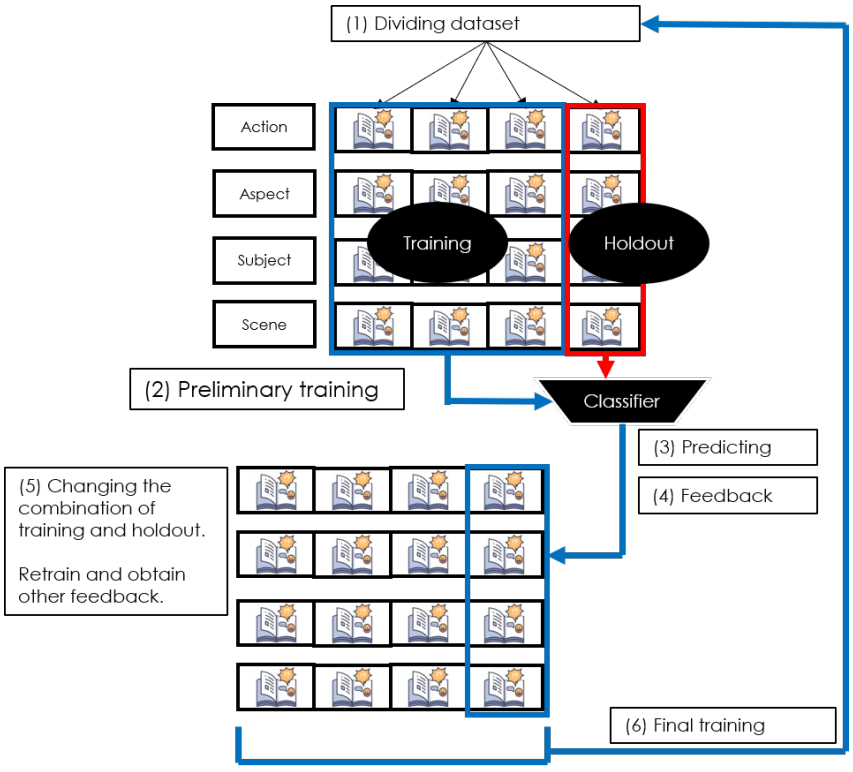
Features	Advantages
■ Transitions 1 Moment-to-Moment 2 Action-to-Action 3 Subject-To-Subject 4 Scene-To-Scene 5 Aspects-To-Aspects 6 Non-sequitur	■ Captured operations on reader's attention

Inter-panel Transitions & Content: Annotators & Base Label Set

- Annotators: Three students.
  1. Annotator 1 & 2: familiar with comic forms
  2. Annotator 3: has experience in creating comics.
- Sets:
  1. Evaluation set:129 panel pairs.
  2. Test set: 2099 panel pairs.
- Each annotator get 1/3 of test set, and the whole evaluation set.

Transitions	Manga109 (2228)
Moment-to-moment	12.6%
Action-to-Action	33.2%
Subject-to-subject	20.4%
Scene-to-scene	10.1%
Aspect-to-aspect	8.3%
Non-sequitur	15.1%

Inter-panel Transitions & Content: Model



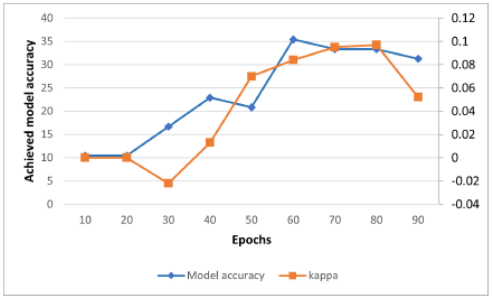
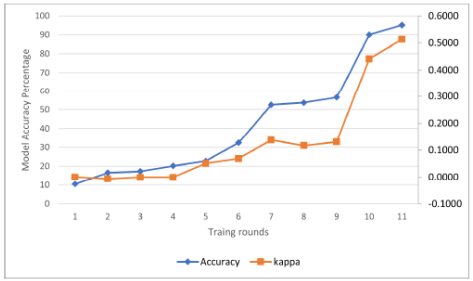
- Two divisions of data:
  1. Training set: trained on training set.
  2. Holding set: tested trained model.
- Feedback by annotators.
- Changed divisions in each rounds.

Inter-panel Transitions & Content: Inter-rater Reliability

- Agreement between annotators
  1. Cohen's kappa: To measure inter-rater reliability for categorical items.
  2. Values: ≤ 0 = no agreement; 0.01–0.20 =slight; 0.21–0.40 = fair; 0.41– 0.60 = moderate; 0.61–0.80 = substantial; 0.81–1.00 = almost perfect agreement.
  3. Result:

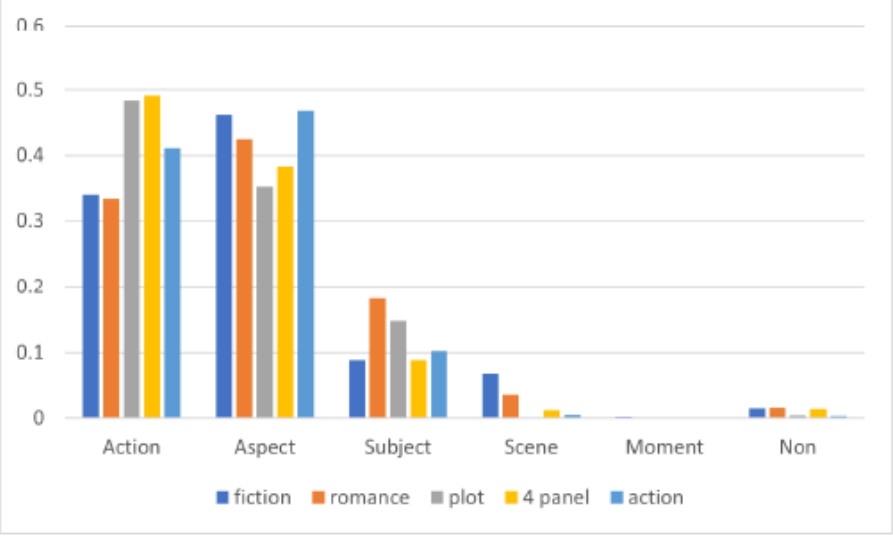
–	Annotator 1&2	Annotator 2&3	Annotator 1&3
reliability (Kappa score)	0.524	0.631	0.774

- Agreement between model and annotator:

CNN, more epochs	Layered CNN, more rounds
	

Inter-panel Transitions & Content:

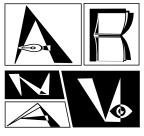
- Genres: Fiction, Romance, Plot, 4 panels, and Action.
- Action transitions and Aspect transitions are most used.
- Romance and Fiction genres used more Aspect transitions; Plot and 4 Panel used more Action transitions.



Intra-panel Relations & Graphical Style

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

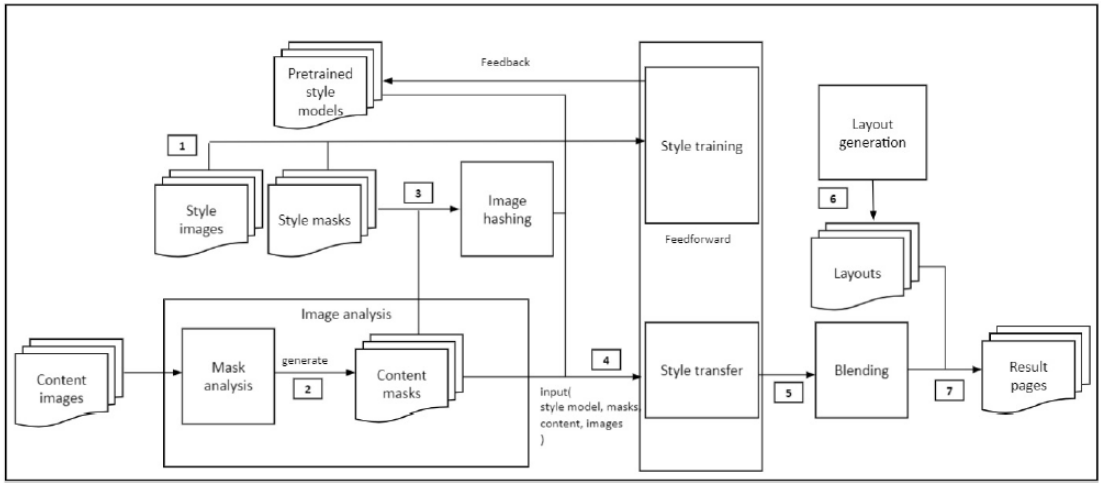
- Corresponding research questions
  - What features influence comic understanding?
  - What are the elements of a comic comprehension model?
  - Can semantics be preserved when drawing style changed?



Intra-panel Relations & Graphical Style: Method and Results

Features	Contributions
■ Feed-forward style transfer with/without perceptual masks ■ Image composition analysis	■ Tested the extent to semantic preserved after drawing style changed ■ Set the framework for style transferring to apply masks

Intra-panel Relations & Graphical Style: Framework

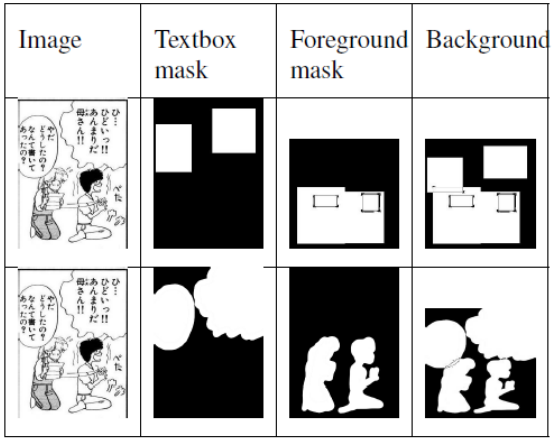


1. Destination style images and masks are used for training.
2. Input images are masked.
3. Inputs' masks and destination images' masks are hashed, and styles are selected by composition similarity.
4. Style transfer module transfers channels in parallel.
5. Per-channel outputs are blended to form a single image.
6. Generate comic layout.
7. Output images are stored and combined with layout.

Intra-panel Relations & Graphical Style: Masks and Compositions

Type of masks:

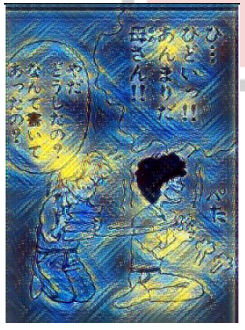
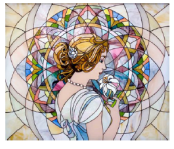
- Rectangle and Fit masks.
- Masks categories: textbox, foreground, background.



Composition selection:

- Combine masks.
- Image hashing: average hash.

Intra-panel Relations & Graphical Style: Style Transfer Results

–	Content	Style	Result
painting, no_mask			
painting, masked		 	

Intra-panel Relations & Graphical Style: Style Transfer Results

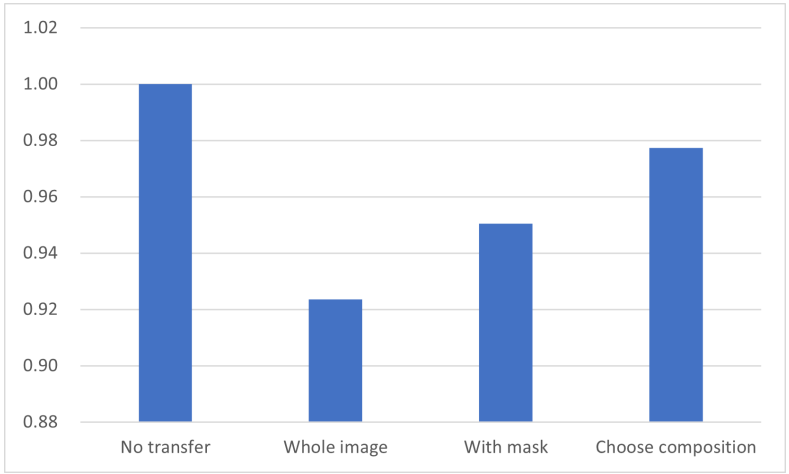
–	Content	Style	Result
comic, no_mask			
comic, masked			

Intra-panel Relations & Graphical Style: Cloze-test Result

Cloze-test.

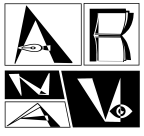
- No style transfer
- Transfer on whole image
- Transfer with masks
- Transfer with composition selection

Compare with no style transfer version, the accuracy be reached.



Comic Generation

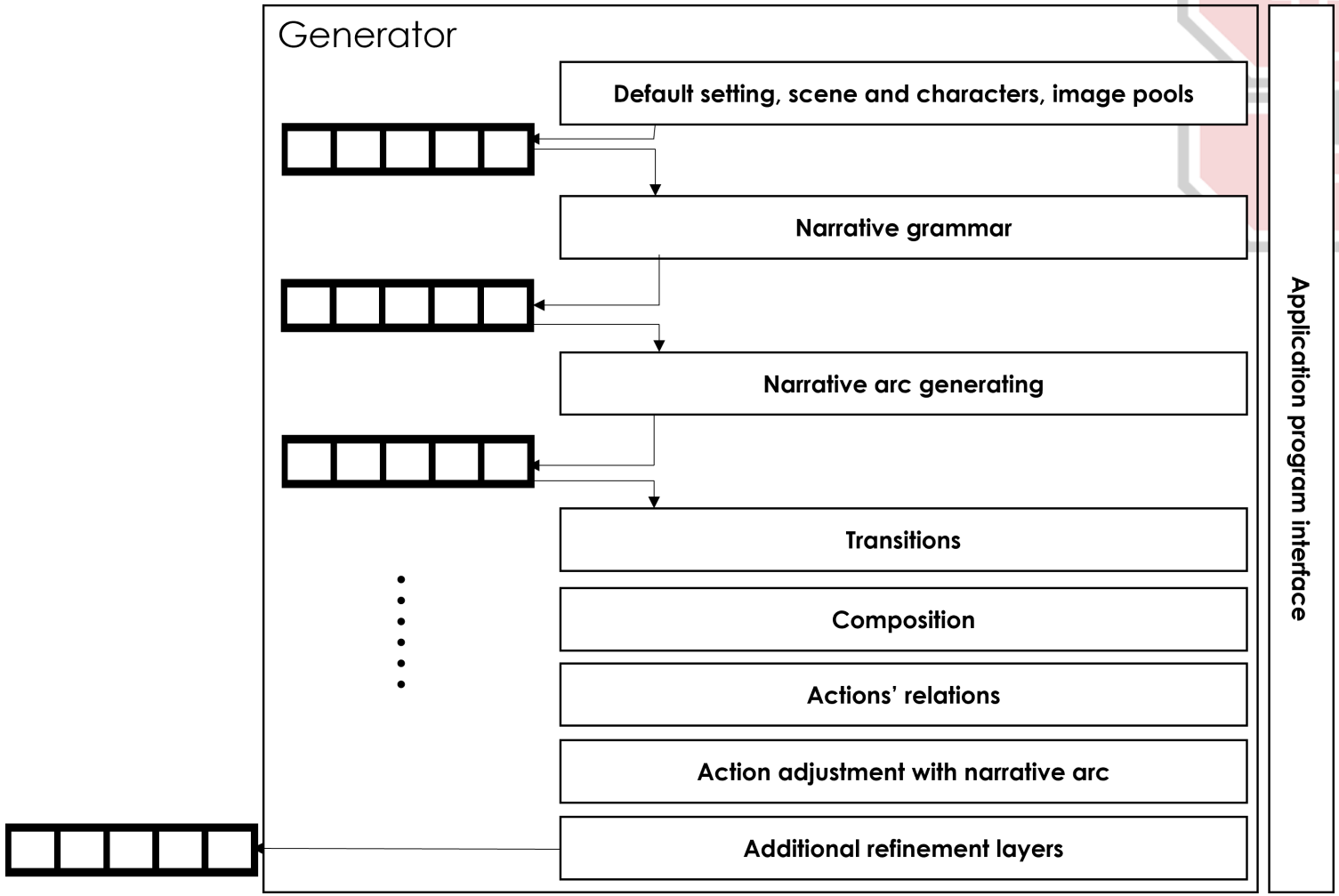
- Motivation
  - An application of comic comprehension model
  - To find comic sequences that match comic theories by human analysis could be high-cost
  - Integrate analyzed components of comics
- Corresponding research questions
  - How comic theories model comic content?
  - Does generated comics form understandable sequences?
  - How comic elements interact?



Comic Generation: Information

Features	Contributions
■ Adjustable generator ■ Application for comic theories ■ Abstract comic metaphors	■ Provided model and API ■ Provided a graphical set for comic generation ■ Implemented comic theories ■ Provided potential test data for comic understanding

Comic Generation: Generator

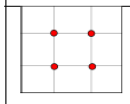
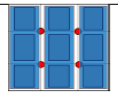
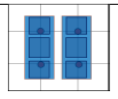
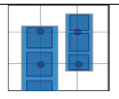
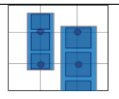
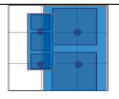
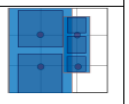


Comic Generation: Graphical Representation

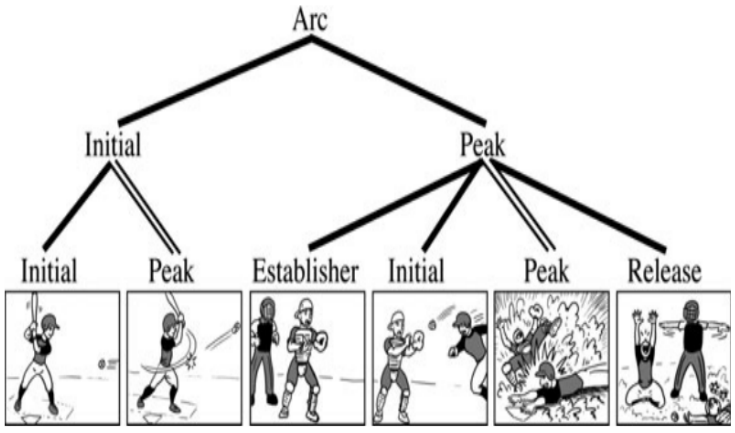
- Abstract symbols:

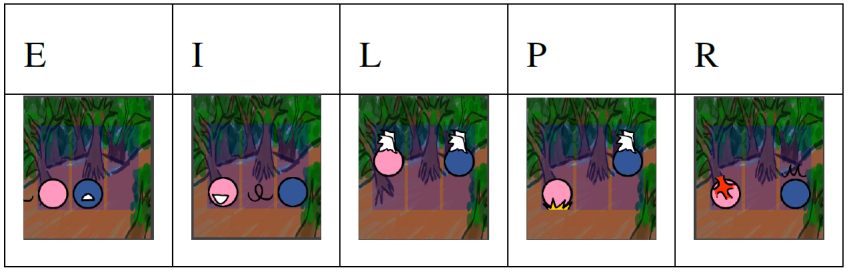
Anger	Quick moving	Slow moving	Anxious	Collision	Relieved	Shock	Big shock
							

- Image compositions:
  1. 3 parallel columns with height levels.
  2. 4 points based on Rule of third.

Rule of third	Basic	Parallel view	Left view with medium shot	Right view with medium shot	Left view with close shot	Right view with close shot
						

Comic Generation: Narrative Structure (Cohn, 2013)





Features

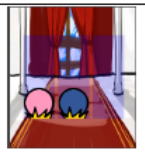
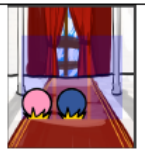
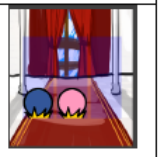
- Categories
  1. Establisher (E)
  2. Initial (I)
  3. Prolongation (L)
  4. Peak (P)
  5. Release (R)
- Phase
  (E)-(I(L))-P-(R)
- Central-embedding

Advantages

- Structuralized narrative

Comic Generation: Local Refinements

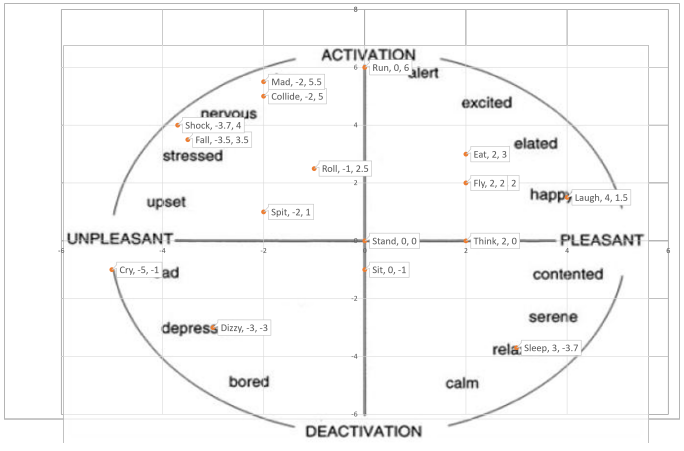
- Panel transitions

Action Pre	Action After	Aspect Pre	Aspect after	Scene Pre	Scene after	Moment Pre	Moment After
							

- Action Network and tension mapping

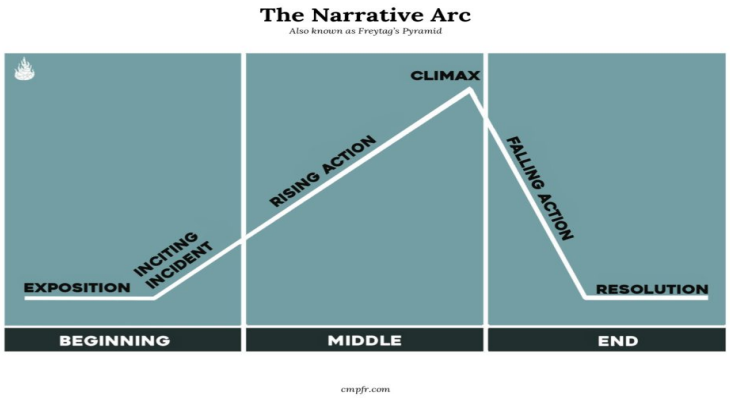
1. Causal relation graph for actions

2. Related action to Circumplex model of affect

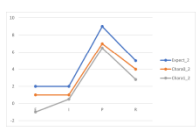


Comic Generation: Narrative Arc & Action Selection

Freytag's Pyramid



Select actions with narrative arc projection

E	L	P	R	narra-arc
				

Features

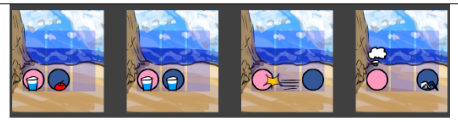
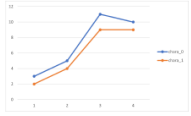
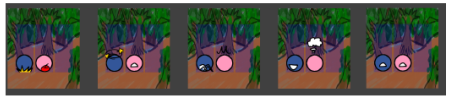
- Categories
  1. Exposition
  2. Rising Action
  3. Climax
  4. Falling Action
  5. Resolution
- Projected to scale [1, 10]
- Select action by tension.
- Possibility:

$$\frac{1}{distance\left(\frac{|s_{i+1}-s_i|^2}{s_{i+1}-s_i}, a_j\right)} \Big/ \sum \frac{1}{distance}.$$

  1. s_i: expected score of panel i
  2. a_j: tension change by action j.

Comic Generation: Results

With/without narrative arc

Settings	Layer: Narrative Grammar and Narrative Arc	Narrative Trend
with		
without		

With/without action relations

Settings	Layer: Action Relations Network Layer
with	
without	

With/without transitions

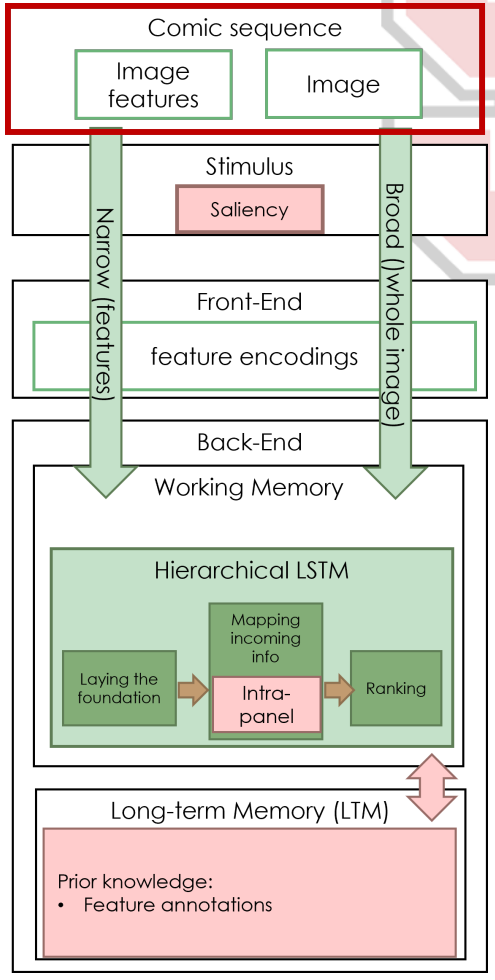
Settings	Layer: Transition layer
with	
transition se-quence	[Aspect-to-Aspect, Action-to-Action, Scene-to-Scene]
without	

Extended layers

Settings	
Textbox layer	
Display layer	

Design Inspiration

- Use comprehension model as a tool to test the results of generating.
  - Enhance content awareness for comprehension model through narrative helping visual features such as transitions, layout, etc.
  - Employ formalized cognitive model to approach the human cognitive process.
  - Interact comic comprehension process with the comic generating process.



Stages

Stage 1: Prototype (confirm the process)

- Construct model frame according to proposed model
- Expected tasks:
  1. Design the interactions and interfaces between modules
  2. Implement and combine the model frame with previous model
  3. Test the process with closure tasks

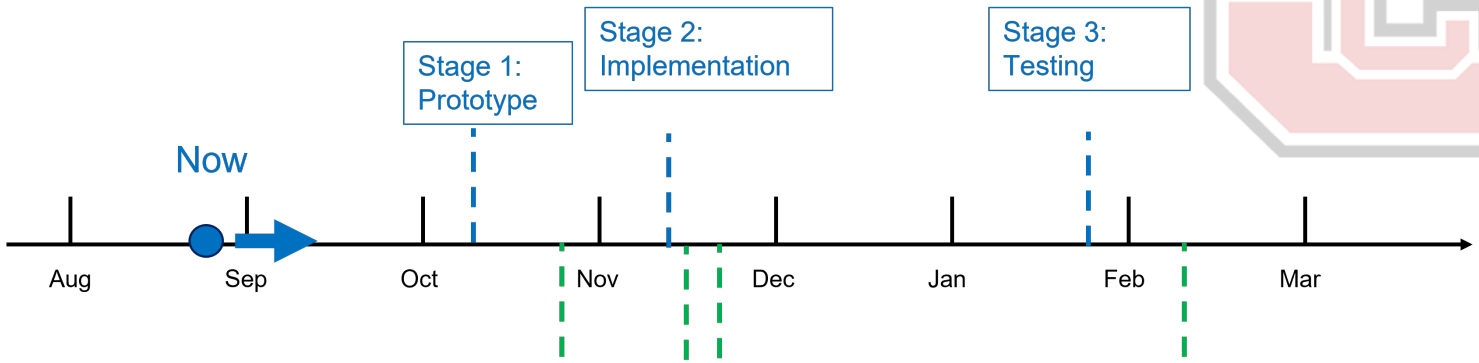
Stage 2: Implementation

- Demonstrate the hypothesis
- Expected tasks:
  1. Integrate encoders in Front-end, feature information into LSTM.

Stage 3: Testing

- Test and compare
- Expected tasks:
  1. Modify comic generator to produce fake cases.
  2. Test generated comics with our model from stage 2.
  3. Compare the results with/ without features.

Timeline



Possible paper submissions for current topic:

NAACL
CVPR
PAKDD
CogSci

Deliverable for side projects (optional):

IUI:
Cooperative authoring interface for comic generating.

Thank you!

Q & A

References I

- Brehmer, M., Lee, B., Bach, B., Riche, N. H., & Munzner, T. (2016). Timelines revisited: A design space and considerations for expressive storytelling. **IEEE transactions on visualization and computer graphics**, **23**(9), 2151–2164.
- Lankow, J., Ritchie, J., & Crooks, R. (2012). **Infographics: The power of visual storytelling**. John Wiley & Sons.
- 3M_Meeting_Network. (1998). Polishing your presentation. https://3rd-force.org/pubs/meetingguide_pres.pdf
- Lopes, P. (2006). Culture and stigma: Popular culture and the case of comic books. **Sociological Forum**, **21**(3), 387–414.
- McCloud, S., & Martin, M. (1993). **Understanding comics: The invisible art** (Vol. 106). Kitchen sink press Northampton, MA.
- Cohn, N. (2013). Visual narrative structure. **Cognitive science**, **37**(3), 413–452.
- Loschky, L. C. et al. (2020). The scene perception & event comprehension theory (spect) applied to visual narratives. **Topics in cognitive science**, **12**(1), 311–351.

References II

- Cohn, N. (2019). Your brain on comics: A cognitive model of visual narrative comprehension.. **Topics in cognitive science**.
- Martens, C., EDU, N., Cardona-Rivera, R. E., & EDU, U. (2020). The visual narrative engine: A computational model of the visual narrative parallel architecture. **8th Annual Conference on Advances in Cognitive Systems**.
- Iyyer, M., Manjunatha, V., Guha, A., Vyas, Y., Boyd-Graber, J., Daume, H., & Davis, L. S. (2017). The amazing mysteries of the gutter: Drawing inferences between panels in comic book narratives. **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**, 7186–7195.
- Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., & Aizawa, K. (2017). Sketch-based manga retrieval using manga109 dataset. **Multimedia Tools and Applications**, **76**(20), 21811–21838.