Composition Proposal Generation for Manga Creation Support

Hironori ITO†∗, Nonmember and Yasuhito ASANO†∗∗, Member

SUMMARY In recent years, cognition and use of manga pervade, and people who use manga for various purposes such as entertainment, study, marketing are increasing more and more. However, when people who do not specialize in it create it for these purposes, they can write plots expressing what they want to convey but the technique of the composition which arranges elements in manga such as characters or balloons corresponding to the plot create obstacles to using its merits for comprehensibility based on high flexibility of its expression. Therefore, we consider that support of this composition technique is necessary for amateurs to use manga while taking advantage of its benefits. We propose a method of generating composition proposal to support manga creation by amateurs. For the method, we also define new manga metadata model which summarize and extend metadata models by earlier studies. It represents the composition and the plot in manga. We apply a neural machine translation mechanism for learning the relation between the composition and the plot. It considers that the plot annotation is the source of the composition annotation that is the target, and learns from the annotation dataset based on the metadata model. We conducted experiments to evaluate how the composition proposal generated by our method helps amateur manga creation, and demonstrated that it is useful.

key words: manga, creation support, composition generation, metadata model, comic engineering

1. Introduction

In recent years, manga comic books have entertained not only children but also people of different generations. Marketers and teachers want to apply manga increasingly as advertisement manga or learning manga recently because manga has benefits of comprehensibility based on high flexibility of its expression. When people who do not specialize in manga creation create it for these purposes, they can write plots expressing what they want to convey, but the technique of composition to arrange elements in manga such as characters and balloons corresponding to the plot presents important obstacles. There are many earlier studies about manga such as image recognition [1], [2], manga retrieval [3], [4], manga classification [5], and comic optimization [6], [7], but there is few studies for helping to draw manga based on what they want to convey. Therefore, we consider that support of this composition technique is necessary for amateurs to use manga while taking advantage of its benefits. In order to support, a method of generating composition proposal from the plot is necessary, so we propose the method using deep learning based on the manga metadata model which we defined newly. Our manga metadata model which we call “M3” expresses elements of plots composed of each action depicted in manga and the composition expressing them. We applied the neural machine translation (NMT) model [8] to this problem because they are the same: they need to predict appropriate labels of variable length from input labels of variable length. In our method, we train the model using the annotation dataset based on M3. This trained model outputs a suitable composition annotation for the plot annotation of input which content that a user wants to express in a frame is expressed in a specified form. Through several experiments, we demonstrated our method for generating the composition proposal is helpful for amateur manga creation.

The remainder of this paper is organized as follows. In Sect. 2, we discuss related studies of the manga metadata model and a system for supporting manga creators. Section 3 explains our manga metadata model (M3) for our method and describes details of the annotation dataset based on the M3. Section 4 proposes the method of composition proposal generation using the NMT model, and explanation of the evaluation experiments and their results are described in Sect. 5. Finally, we summarize the overall research in Sect. 6.

2. Related Work

Firstly, we mention a research for supporting manga creators by Ueno et al. [9]. They aimed at automatic creation of manga pictures, and tried to propose a system which can create pictures flexibly based on objects in the prepared dataset. Their work focuses on the set of pictures which contains some kinds of object transition. Their method generates the second pictures based on the first picture a user added changes automatically with utilizing picture objects transition database. In contrast to this research, our research proposes a method for the optimization of the objects arrangement based on the action that creators want to express after defining what they call object by the metadata model.

Secondly, we mention about 2 earlier studies of the manga metadata model have been conducted by Morozumi et al. and by Fujimoto et al. Morozumi et al. [10] proposed a manga metadata
framework for discovering and accessing each element in manga, and reusing them for manga creation. This framework is contrived from three viewpoints: bibliographic description of manga as published material; structure description on manga expressions and stories; and ontology description on concepts, major characters, and subjects appearing in manga. This is a suitable model for reusing elements created before in manga creation and searching elements in manga from the viewpoints. When we define the part of entity in M3, we used some entities define in Morozumi’s model as reference.

Then, Fujimoto et al. [11] aimed to develop manga searching and recognition technologies for characters and texts in manga. Then, they constructed an open academically-available, large-scale, and high-quality manga dataset. They proposed a metadata model used for annotation of their dataset. The metadata model by Fujimoto et al. is defined from three viewpoints of frame, character, and text, which are given shape and area by a rectangular enclosure. This simple model consists only of spatial coordinates, where each element exists in manga, the name, face, and whole body of the character, and the contents of the dialogue. This model is presented in Fig. 1.

Their models can not create dataset which has composition elements, plot elements and relational between them. The dataset is necessary for our proposed method. Therefore, M3 can describe the relation between expressions in manga and actions in the plot which express the content. We define each element in M3 based on the definitions above. The overall of M3 is presented in Fig. 2.

For this research, we define “plot” as consisting of multiple actions described from the viewpoint of the subject, target, motion, and emotion based on “How to write a plot of manga”[13] and “composition” as what indicates how to express characters or balloons in a frame from perspectives of angle, object size, and placement based on “How to draw Illustration and Manga”[14]. We use actions for expressing the plot because it is crucially important to conveying the story in manga and because it is closely tied to the composition.

We define each element in M3 based on the definitions above. The overall of M3 is presented in Fig. 2.

In Fig. 2, we define the entity set surrounded by the red dotted line as the composition metadata model and the entity set surrounded by the green dotted line as the plot metadata model. In addition, the entities surrounded by the blue solid line are entities not defined in the earlier manga metadata model, which are newly defined to express the plot and the important element for the composition. We mainly explain the newly defined elements in this paper.

3. Manga Metadata Model for Composition Recommendation

This section introduces a manga metadata model (M3) to apply machine learning to manga from the viewpoint of plot and composition. This model is the foundation for our proposed method because it is utilized as an infrastructure for constructing annotation dataset used in our method. We define the necessary elements for expressing the composition and the plot including their relevance. Then we feed back experimentally results obtained from an earlier study [12] to the model.

For this research, we define “plot” as consisting of multiple actions described from the viewpoint of the subject, target, motion, and emotion based on “How to write a plot of manga”[13] and “composition” as what indicates how to express characters or balloons in a frame from perspectives of angle, object size, and placement based on “How to draw Illustration and Manga”[14]. We use actions for expressing the plot because it is crucially important to conveying the story in manga and because it is closely tied to the composition.

We define each element in M3 based on the definitions above. The overall of M3 is presented in Fig. 2.

In Fig. 2, we define the entity set surrounded by the red dotted line as the composition metadata model and the entity set surrounded by the green dotted line as the plot metadata model. In addition, the entities surrounded by the blue solid line are entities not defined in the earlier manga metadata model, which are newly defined to express the plot and the important element for the composition. We mainly explain the newly defined elements in this paper.

3.1 Composition Metadata Model

First, we explain the structure and details of each new part in the composition metadata model. In the composition metadata model, the title entity expressing information about a title of manga including the volume number is the top entity of the structure. The page entity showing page number in the manga links the title entity. Also, the frame entity indicating the order number to read the frame in the page links the page entity. The frame is the smallest unit as a scene constituting manga. Therefore, it is used as a basic unit for combining plot and composition in the proposed metadata model. This frame entity has only information related to reading order on each page. Other information is embodied by the following 11 entities.

The area entity has information related to the coordinates of the frame within the page. The grid-line entity has information about whether the grid-line of the frame exists. Although it is not defined as an entity in earlier studies, the existence of a grid-line is an important element that is largely involved in composition. For that reason, we add it into our metadata model.

The background entity has information about places...
depicted in the frame. An entity about the place as the stage through the manga is included in Morozumi’s model, but the place depicted as the background in each frame changes from frame to frame depending on the story. Therefore, we define it as an entity linked to the frame entity.

The scene type entity expresses the object of focus in the frame from three kinds of “Character,” “Landscape,” and “Others.” Whether particularly addressing “Character” or “Landscape” in the composition is a factor that is closely related to the story. For that reason, we added it into M3.

The angle entity expresses the angle from which the focused object in the frame is drawn. This entity is represented by three labels: “Swing,” “Horizontal,” and “Overhead.” The change of angle is necessary for preventing an unnatural flow of frames because of the turning back of the character’s standing position and for directing the flow of natural camera work along the story. Therefore, it is an important element in the composition from the viewpoint of comprehensibility of the story.

The distance entity expresses the distance to those included in the frame from two types of “Up” and “Long.” We define the composition drawing the focused object from near as “Up” and the composition drawing the focused object from far away as “Long.” To convey the psychological description of the character, it is effective to distinguish the mode of drawing “Up” or “Long” as the composition. It is an element closely related to the story, so we incorporate it into our metadata model.

The character entity has information for each character in the frame, and is embodied by four entities of “Character ID,” “Area,” “Direction,” and “Body Part.” The character ID entity expresses a character ID uniquely assigned to the main character for each manga. The area entity has information about the coordinates of the character. The direction entity expresses the body direction of the character by “Front” or “Back.” The body part entity expresses the character part depicted in the frame by five kinds of “Whole body,” “Only upper body,” “Only face,” “Only lower body,” and “A part of body.” The character ID and area entities are core elements for expressing a story. The direction and body part entities are also important elements for expressing the composition. For that reason, they are defined as entities expressing the character.

The balloon entity, which includes information about each speech balloon in the frame, is embodied by four entities of “Speaker,” “Area,” “Shape,” and “Contents.” The speaker entity expresses the speaker of the balloon by the character ID above. The area entity has information about its coordinates. The shape entity expresses the shape of the balloon. The content entity expresses the letters written in the balloon. This is also defined in earlier studies, and it is noted in them that the balloon entity is indispensable in progress of the story as with the character entity. In our metadata model, we newly define the speaker and the shape entities because it is important to express the composition and to convey the story effectively.

Onomatopoeia is an expression drawn by designing onomatopoeic and mimetic words. The onomatopoeia entity has information about each onomatopoeia in the frame, and is embodied by two entities of “Notation” and “Area.” The notation entity expresses the notation of the onomatopoeia. The area entity has information about its coordinates. The reason we define it as an entity is that it is an important directly to convey a situation as a sensation.

The prop entity has information about each tool used or touched by a character in the frame. It comprises three entities of “Name,” “Area,” and “User.” The name entity expresses the tool name. The area entity has information about its coordinates. The user entity expresses the character ID of one who uses or touches the tool. Morozumi states that props such as belongings of characters should have an important meaning in the story. Therefore, we incorporate it into M3.

3.2 Plot Metadata Model

Next, we explain the structure and details of each part in the plot metadata model. We consider that the place, the subject, the emotion, the target, and the verb are the basis for expressing story based on “How to write a plot of manga.” We define the plot metadata model using the following entities.

Like the composition metadata model, the top-level entity in the plot metadata model is the title entity. The scene entity is defined for expressing a switch of place in the manga. It is embodied by two entities: the action entity indicating each action performed in the scene and the place entity indicating the place where the action is performed in the scene. Some definitions of scene in manga are given [15], but we define a scene as a set of actions performed in each place. Therefore, the place entity has information related to the name of the place where the actions are performed in each scene. This is one for each scene.

The action entity has information of each character’s action. It is embodied by five entities: “Subject,” “Verb,” “Emotion,” “Target,” and “Content.” The subject entity expresses the principle of the action, and it is noted by the character ID. The verb entity expresses a verb representing the action. The emotion entity using the expressions of 16 kinds proposed by Ekman et al. [16] expresses the feelings of the subject of the action such as “Anger,” “Sadness,” “Fear,” “Surprise,” and so on. The target entity expresses the target on which the action works, using the character ID or the prop. The content entity expresses the letters, when the action is accompanied with words such as an utterance or thought by a character.

The composition proposal generation using machine learning requires plot annotation describing the contents expressed in a manga frame structurally, but there was no dataset collected from the viewpoint. It also requires composition annotations describing each element, such as characters depicted in a frame, which link to the plot annotation. Therefore, we constructed an annotation dataset based on
M3. This is one of our contributions.

In constructing the dataset, we selected three comics that were made into mangas based on Japanese classical literature: “Botchan” [17], “The Night of the Milky Way Train” [18], and “The Crab Cannery Ship Author” [19]. The number of each element in the dataset is shown in Table 1. “Onoma” in the table means the onomatopeia.

4. Manga Composition Proposal Generation

In this section, we propose a method of manga composition proposal generation. The method uses deep learning trained with an annotation dataset that we introduced in the previous section. The problem of manga composition generation can be regarded as a problem of receiving a multilabel of variable length as input and predicting and outputting a variable length multilabel corresponding to it. Tasks of machine translation pose similar problems. We consider that we can solve the problem by giving a plot annotation as a translation source sentence and composition annotation as a target sentence similarly.

For use in our research, we propose a composition proposal generation using Attention2D [8], which is an NMT model. The flow of the composition recommendation method is shown in Fig. 3.

We explain the details of input and output of the proposed method in Sect. 4.1, and describe the outline of Attention2D in Sect. 4.2. Then, we conducted several preprocessings for the purpose of making it possible to use an annotation data of multi manga cross-sectionally for model training and making the model easy to learn by consolidating data and reducing types of labels. We explain its detail in Sect. 4.3.

4.1 Input and Output

In the proposed method, a list of labels which expresses the content of a frame is given as input. It shows one action written according to the action entity. The number of list increases according to the number of actions expressing the contents in a frame. It is represented in the same form as the action entity like [“Subject”, “Verb”, “Target”, “Emotion”, “Content”]. For example, when describing the content such as “Chara2 gets angry and hits Chara1 while saying somewhat,” in a frame, its input is represented as [[Chara2, hit, Chara1, anger, somewhat], [Chara1, hit, Chara2, anger, ]].

In the output, the composition annotation is generated as expressed in accordance with each entity of Area, Grid-Line, Scene Type, Angle, Distance, Character, Balloon, Onomatopeia, and Prop in the composition metadata model, according to the input. The entities related to the whole frame of Area, Grid-Line, Scene Type, Angle, and Distance invariably include one entry for every input. The number of each entity of Character, Balloon, Onomatopeia, and Prop changes according to the input. This output samples are shown in Fig. 4 and Fig. 5. In the samples, SX and SY mean the coordinates which indicate the area start point expressed by the distance from the upper left corner of the frame. Additionary, width, height, SX, and SY are designated in pixels.

4.2 Attention2D

Our method is realized by the NMT model. The most famous one in NMT is Seq2Seq [20], and we use Attention2D model that implements this Encoder-Decoder model using a development of Convolutional Neural Networks [21] called DenseNet [22]. In this section, we explain why we use Attention2D for manga composition recommendation.
There are two reasons. The first reason is that information related to the appearance position of the label can be incorporated as an input in this model. In our proposed method, the relation between plot annotation and composition annotation is trained on a frame unit basis. Then, the order is fixed for each action in the source and each composition element such as the frame and character in the target. Therefore, we consider that we can achieve higher prediction accuracy using Attention2D.

The second reason is that Attention2D can use the attention value corresponding to the previous output when generating each label. In attention mechanisms such as Seq2Seq [20] developed to date, it was common that calculation of the attention value has already been completed before the generation and this value is used as input to the decoder irrespective of the generated labels. Nevertheless, it is difficult to preserve the naturalness of the whole by this mechanism because it cannot devote attention to highly relevant words after considering the words generated so far. Attention2D, however, uses attention values considering the generated labels, so it is easier to generate natural output overall than ever to. In composition annotation, which is the output of our method, for example, the character size changes according to the size of the outputted frame. The balloon position changes according to the character position. For that reason, strong correlation exists between units. The feature is suitable for Attention2D.

Then, it is necessary to set several hyper parameters for Attention2D. Therefore, we set embedding size = 128, number of layers = 24, and growth rate = 3 in our method, which archived the highest prediction accuracy in searching hyper parameters by Elbayad et al [8].

4.3 Preprocessing

Some difficulties arise when using the annotation dataset as it is. For example, training does not advance well because there are labels of too many kinds for the dataset. Therefore, we need a method to prepare multiple comic annotation data cross-sectionally for model training and to advance training well by concentrating data and reducing types of labels.

First, as for the point that the character ID varies according to manga, which is the most problematic when treating the annotation data for plurality of manga crosswise, we assigned importance to the character IDs in each manga based on the appearance frequency of the character in the plot annotation data of each manga, and relabeled the ID based on it. Therefore, we made it possible to handle character information across multiple manga.

We also need to proceed with training well with a small amount of data. Then, we vectorize verbs indicating actions in plot annotation data using Word2Vec [23] that have been trained by Wikipedia, and classify the vectorized verbs into 20 classes using Kmeans and label them according to the class because the variation of verb is too wide ranging.

Additionally, we classify every three letters according to the number of letters for entities including letters such as the balloon and the onomatopoeia. Then, we label them according to the class. This is why the number of letter is more importance than its content in the composition.

Then, as for each element peculiar to the composition annotation data such as area, we consolidate labels by labeling them every 25 pixels. Through this preprocessing, we make it possible to apply a neural machine translation model to the plot and composition annotation dataset.

5. Evaluation

This section presents an explanation of the quantitative and qualitative evaluation performed on our proposed method. First, we conducted accuracy evaluation using the word error rate (WER) for quantitative evaluation. We explain this details and results in Sect. 5.1. In addition to confirming whether our proposed method can support the amateur’s manga creation, we asked 4 subjects to create manga frames using comiPo! [24] based on actually generated composition annotations from the plot annotations. Then, we asked another 6 subjects to evaluate them from the viewpoints of ease of understanding, naturalness and comparison between frames. We explain the details and the obtained results in Sect. 5.2.

5.1 Quantitative Evaluation

We evaluated the accuracy using WER to confirm the training of the Attention2D model using an annotation dataset. The WER is an evaluation index often used for machine translation. This is defined as follows.

$$WER = \frac{S + D + I}{N}$$  \hspace{1cm} (1)

In the above equation, $S$ is the number of substitutions, $D$ is the number of deletions, $I$ is the number of insertions, and $N$ is the number of words in the reference.

For this evaluation, we used 100 frames not used for training as test data. We calculated WER for all elements.
related to the composition in each frame and for each element. Then we obtained their averages. The results are presented in Table 2. “Chara”, “Ball”, and “Ono” in Table 2 represent the results of character, balloon, and onomatopoeia entities respectively.

Although the results described above are higher than WER by recent neural machine translation, we consider that the accuracy reaches a practical level of usefulness. We infer that the shortage of the dataset is the main reason why it produced higher values than the WER by the recent neural machine translation. As a dataset commonly used in neural machine translation, WMT’14 consists of about 5 million pairs of sentences. In the field of neural machine translation, datasets containing greater quantities are often used. These datasets contain more words, including unknown words. By contrast, our dataset has few pairs, but only a fixed number of labels as the number of words. It is also made as small as possible in the preprocessing of our method. Therefore, the WER for our method did not increase greatly.

When analyzing the actual mispredictions in detail, no mistake was found in terms of the position or number of labels of each element. Then, we infer that the WER will be improved by increasing the dataset because many mispredictions occurred within the label of the same attribute. In addition, the WER for the character was the worst, but this derives from the fact that someone repeatedly generated a list indicating the same character in the output up to the upper limit. In this model, although the upper limit of the generated label for the whole was incorporated. There was no upper limit for element units. Therefore, this point is expected to be improved by providing this mechanism.

5.2 Quality Evaluation

We conducted the following qualitative evaluation to evaluate how much we can support manga creation by creators, even amateurs with our proposed method.

For this evaluation, we asked 4 subjects consisting of 2 professionals (A, B) and 2 amateurs (C, D) to create based on the annotations prepared for each frame. The 30 frames extracted from 3 comics included in the annotation dataset are used for this experiment. Then, they are divided into 2 groups X and Y of 15 frames each. We divided creators into 2 teams: Team 1 consisting of A and C, and Team 2 consisting of B and D. Team 1 has the situation explanation for the frames in the group X and their plot annotations and also has the composition annotations generated by the proposed method from the plot annotations in addition to them in case of the frames in the group Y. However, Team 2 has the set of swapped ones for the frames in the group X and group Y. The samples of plot annotation for the frames are shown in Figs. 6 and 7.

They created the frame with imagination only based on them if the situation explanation and the plot annotation of the frame are only given. They created the frame according to it fundamentally if the composition annotation of the frame is given in addition to them. However, we allowed them to modify, delete, and insert elements in the composition annotation if it becomes an obfuscating expression after compliance with it because the objective of this proposed method is support of manga creation.

Next, we asked the other 6 subjects for 3 evaluations to the total 120 frames created as described above. In these evaluations, they do not know whether the author of the frame is a professional or an amateur. In the first evaluation, they evaluated each created frame in 3 stages of Good, Neither, or Bad whether the content could be expressed in an easy-to-understand manner than the situation explanation. In the second evaluation, they evaluated whether or not each created frame was natural using three stages of responses: Natural, Neither, and Unnatural. In the third evaluation, they compared the frames expressing the same situation by different authors. They chose which is better considering clarity and naturalness comprehensively.

We present results of the first experiment in Fig. 8 and those of the second experiment in Fig. 9. These charts show the number of evaluations acquired by the frames, as created by each author. We present the result of the third experiment in Fig. 10. The chart shows the number chosen as better when comparing two frames. In these charts, the notation with prime means the frames created based on the composition recommendation. The notation without prime represents the frames created without composition recommendation.

From Fig. 8, in the case of professionals A and B, even for the frames created based on the composition proposal, the same evaluation as that created without it was obtained. In the case of amateurs C and D, the frames created based on the composition proposal got higher evaluation for understandability. We present two examples in Fig. 11 and Fig. 12. These figures express the same situation that a friend stands up unexpectedly in front of the main character. The plot is as shown in Fig. 6. Its composition proposal is as shown in Fig. 4 and this sample follows the proposal faithfully. Figure 11 is created by the amateur C with the composition proposal and acquired many good evaluations in the first evaluations. Figure 12 is created by the amateur D without the composition proposal and acquired many
bad evaluations in the first evaluation. We consider that the evaluation of Fig. 11 was good because of the onomatopoeia placement and character placement, while the evaluation of Fig. 12 was poor because the character’s scale is incorrect and it is difficult to see if it stood up now.

Although the frames created based on the composition proposal obtained slightly lower evaluation than the frames created without it in the case of professionals, we found the same tendency for naturalness for amateurs from Fig. 9. We present two examples in Fig. 13 and Fig. 14. These figures express the same situation that a employee is accused of a big mistake by the president and the president says “It is no wonder that you’re fired. Didn’t I send you to avoid this situation? Do you know some of our losses from that strike? Asakawa?”. The plot is as shown in Fig. 7. Its composition proposal is as shown in Fig. 5 and this sample follows the proposal faithfully. Figure 13 is created by the amateur C with the composition proposal and acquired many good evaluations in the second evaluation. Figure 14 is created by the amateur D without the composition proposal and acquired many bad evaluations in the second evaluation. We consider that the evaluation of Fig. 13 was good because of the division of the balloons and the placement of the character and prop, while the evaluation of Fig. 14 was poor because the too much letters are packed into the balloon and the positional relationship between the characters and the balloons is unnatural.

Then, Fig. 10 shows that, in the case of professionals,
there are many good ones made by themselves without composition proposal when comparing each frame. However, in the case of amateurs, the number selected as better is higher than in the case of creating without it. From the above, although the quality is somewhat inferior to those created by professionals, it is shown to be sufficiently useful as help for amateurs' manga creation.

6. Conclusion

As described in this paper, we proposed the composition proposal generation method with deep learning using the manga annotation dataset based on our newly defined manga metadata model for machine learning. We define the elements expressing composition and plot after making these information available for machine learning in the metadata model. Then, we constructed an annotation dataset based on it. Then, we trained the NMT model for composition proposal generation using the plot annotation indicating the action in each frame of the as an input sentence and the composition annotation corresponding thereto as a pair based on this dataset.

As an evaluation of this composition proposal generation method, we conducted a quantitative evaluation using WER and a qualitative evaluation comparing the frames created based on the composition proposal and the frames created based on the imagination. From the quantitative evaluation, we obtained a practically applicable level of accuracy. Through the qualitative evaluation, we also confirmed that our method achieved our objective of supporting manga creation by amateurs.

Acknowledgments

This research was supported by JSPS KAKENHI Grants-in-Aid for Scientific Research(C) No. 15K00423 and No. 17K01590 and a Grant-in-Aid for Scientific Research on Innovative Areas No. 18H05031. We sincerely appreciate their support.

References

[1] X. Pang, Y. Cao, R.W. Lau, and A.B. Chan, “A robust panel extraction method for manga,” Proc. 22nd ACM international conference on Multimedia, pp.1125–1128, ACM, 2014.
[2] A.K.N. Ho, J.C. Burie, and J.-M. Ogier, “Panel and speech balloon extraction from comic books,” 2012 10th IAPR International Workshop on Document Analysis Systems (DAS), pp.424–428, 2012.
[3] K. Hoashi, C. Ono, D. Ishii, and H. Watanabe, “Automatic preview generation of comic episodes for digitized comic search,” Proc. 19th ACM international conference on Multimedia, pp.1489–1492, ACM, 2011.
[4] Y. Matsui, “Challenge for manga processing: Sketch-based manga retrieval,” Proc. 23rd ACM international conference on Multimedia, pp.661–664, ACM, 2015.
[5] W.-T. Chu and Y.-C. Chao, “Line-based drawing style description for manga classification,” Proc. 22nd ACM international conference on Multimedia, pp.781–784, ACM, 2014.
[6] W.-T. Chu and C.-H. Yu, “Optimized speech balloon placement for automatic comics generation,” Proc. 3rd ACM international workshop on Interactive multimedia on mobile & portable devices, pp.1–6, ACM, 2013.
[7] Y. Cao, A.B. Chan, and R.W.H. Lau, “Automatic stylistic manga layout,” ACM Transactions on Graphics (TOG), vol.31, no.6, pp.141, 2012.
[8] M. Elbayad, L. Besacier, and J. Verbeek, “Pervasive attention:2d convolutional neural networks for sequence-to-sequence prediction,” SIGNLL Conference on Computational Natural Language Learning 2018, 2018.
[9] M. Ueno, N. Mori, and K. Matsumoto, “2-scene comic creating system based on the distribution of picture state transition,” Distributed Computing and Artificial Intelligence, 11th International Conference, vol.290, pp.459–467, Springer, 2014.
[10] A. Morozumi, S. Nomura, M. Nagamori, and S. Sugimoto, “Metadata framework for manga: A multi-paradigm metadata description framework for digital comics,” International Conference on Dublin Core and Metadata Applications, pp.61–70, 2009.
[11] A. Fujimoto, T. Ogawa, K. Yamamoto, Y. Matsui, T. Yamasaki, and K. Aizawa, “Manga109 dataset and creation of metadata,” Proc. 1st International Workshop on coMics ANalysis, Processing and Understanding, pp.1–5, ACM, 2016.
[12] I. Hironori and A. Yasuhito, “Manga metadata model for manga creation supporter,” Proc. 10th Data Engineering and Information Management Forum, Fukui, Japan, E+5, March 2018.
[13] K. Oishi, How to write a plot of Manga, Genshisha, 2012.
[14] C. STUDIO, “How to draw illustration and manga,” 2018.
[15] H. Chie and M. Tetsuya, “Automatic extraction of scene of manga using the attribute of frames,” The 30th Annual Conference of the Japanese Society for Artificial Intelligence, 2016.
[16] P. Ekman and D. Keltner, “Universal facial expressions of emotion,” Segerstrale U, P. Molnar P, eds. Nonverbal communication: Where nature meets culture, pp.27–46, 1997.
[17] S. Natsume (Story) and Team-Bangmicas (Drawing), Botchan, East Press, 2016.
[18] K. Miyazawa (Story) and Variety-Artworks (Drawing), The Crab Cannery Ship Author, East Press, 2007.
[19] T. Kobayashi (Story) and Variety-Artworks (Drawing), The Crab Cannery Ship Author, East Press, 2007.
[20] I. Sutsekey, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” Advances in neural information processing systems, pp.3104–3112, 2014.
[21] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “Imagenet classification with deep convolutional neutral networks,” Advances in neural information processing systems, pp.1097–1105, 2012.
[22] G. Huang, Z. Liu, L. Van Der Maaten, and K.Q. Weinberger, “Densely connected convolutional networks,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.2261–2269, 2017.
[23] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” Advances in neural information processing systems, pp.3111–
Hironori Ito received the B.S. and M.S. degrees in Informatics from Kyoto University in 2017 and 2019, respectively. He is now with Recruit Co., Ltd.

Yasuhito Asano received the BS, MS, and DS degrees in information science from University of Tokyo in 1998, 2000, and 2003 respectively. In 2003–2005, he was a research associate in the Graduate School of Information Sciences, Tohoku University. In 2006–2007, he was an assistant professor in the Department of Information Sciences, Tokyo Denki University. He joined Kyoto University in 2008 as an assistant professor. In 2009–2018, he was an associate professor in the Graduate School of Informatics, Kyoto University. Since 2019, he has been a professor at Faculty of Information Networking for Innovation and Design, Toyo University. His research interests include web mining and network algorithms. He is a member of the IEICE, IPSJ, DBSJ and OR Soc. Japan.