Neural Machine Translation of Artwork Titles Using Iconclass Codes

Nikolay Banar[1, 2] **Walter Daelemans** [1] **Mike Kestemont** [1, 2]
[1]CLiPS, University of Antwerp, Antwerp, Belgium
[2]ACDC, University of Antwerp, Antwerp, Belgium
{nicolae.banari, walter.daelemans, mike.kestemont}@uantwerpen.be

Abstract

We investigate the use of Iconclass in the context of neural machine translation for NL↔EN artwork titles. Iconclass is a widely used iconographic classification system used in the cultural heritage domain to describe and retrieve subjects represented in the visual arts. The resource contains keywords and definitions to encode the presence of objects, people, events and ideas depicted in artworks, such as paintings. We propose a simple concatenation approach that improves the quality of automatically generated title translations for artworks, by leveraging textual information extracted from Iconclass. Our results demonstrate that a neural machine translation system is able to exploit this metadata to boost the translation performance of artwork titles. This technology enables interesting applications of machine learning in resource-scarce domains in the cultural sector.

1 Introduction

In the age of mass-digitization, cultural heritage institutions put significant effort in making their (meta) data available to developers and researchers. Artificial intelligence, and machine learning in particular, increasingly plays an important role in this process (Fiorucci et al., 2020). Recent case studies have demonstrated successful applications of machine learning methods to cultural heritage collections. Most of this work relies on advances in computational methods and utilizes a modelling framework known as deep neural networks (LeCun et al., 2015; Schmidhuber, 2015). However, such algorithms are data-intensive and require large annotated datasets, which recently have become available in some fields (Tiedemann, 2012; Krizhevsky et al., 2012; Lin et al., 2014). These datasets contain millions of training items, which allowed researchers to achieve impressive results in many tasks. However, the construction of such materials in the domain of cultural heritage material is an even more expensive process, as it requires the intervention of highly-trained subject experts. Hence, many institutions can only offer smaller datasets, that contain just a fraction of the number of training examples that are needed to train a deep learning algorithms. Transfer learning is a common solution to overcome such a lack of training data (Ruder et al., 2019). In neural machine translation (NMT), networks are nowadays commonly pre-trained on large generic datasets of parallel sentences, before they get fine-tuned on a more specific "downstream" corpus. Such networks, however, are conventionally only exposed to the actual sentence pairs in the target domain and are ignorant of additional knowledge that might be available such as, for example, iconographic metadata about objects and their relations. In the case of artworks, computational methods that can exploit such additional knowledge are highly appealing.

This work aims to apply NMT in the context of cultural heritage metadata using Iconclass (Vellekoop et al., 1973) as a source of external knowledge. Iconclass (see Section 3.1.1) contains keywords and definitions of subjects represented in artworks. We propose a simple approach to integrate this external knowledge in an NMT architecture for artwork titles to improve the translation performance. The structure of this paper is as follows. We first present the related work in Section 2. Then, we describe the datasets and present the applied methods in more detail in Section 3. Next, we present the results of our

Proceedings of LaTeCH-CLfL 2020, pages 42–51
Barcelona, Spain (Online), December 12, 2020.

case study and discuss them in Section 4. Finally, we summarize our main contributions and findings with proposals for future work in Section 5.

2 Related Work

Modern NMT systems nowadays often work at the level of an individual sentence pair and aim to translate a source sentence into a target sentence, without making use of additional information other than the source sentence itself. The idea to concatenate a source sentence with additional information, however, is not new. This preprocessing step is appealing due to its simplicity and model-agnostic applicability. Previous approaches in this respect are generally divided into 2 categories (see examples in Table 1): (i) *extended context* (Tiedemann and Scherrer, 2017), where additional information in the source language is added to the source sentence (and sometimes to the target sentence); (ii) *data augmentation* (Bulté and Tezcan, 2019), where the source sentence is enriched with information in the target language.

Tiedemann and Scherrer (2017) investigated the benefits of the extended context approach in attention-based NMT for DE→EN subtitles (see Table 1). The source sentence was concatenated with the previous source sentence and, then, the same technique was additionally applied to the target sentence. They used a special prefix to mark tokens belonging to the extended context. Although the improvement over the baseline was moderate, the NMT models were able to utilize the additional context and to distinguish it from the main sentence. In follow-up work, Bawden et al. (2018) designed EN→FR test sets to investigate the usefulness of the previous source and target sentences in the context of NMT. They demonstrated that the concatenation strategy leads to improved performance. Agrawal et al. (2018) applied the concatenation technique with a Transformer-based architecture to EN→IT TED talks and experimentally varied the number of concatenated sentences included. There too, the extended context was demonstrated to be beneficial for Transformers. Junczys-Dowmunt (2019), finally, developed one of the best-performing systems based on the same idea in the context of the WMT19 news translation shared task for EN→DE.

Bulté and Tezcan (2019) proposed a simple and efficient data augmentation method for NMT that yielded substantial performance improvements for EN→NL and EN→HU. The source sentence was concatenated with fuzzy matches, or sentences in the target language retrieved from a translation memory, that covered the entire training set. The fuzzy matches were selected on the basis of a simple similarity measurement between each source sentence and all other source sentences from the translation memory. Then, the fuzzy source sentences with a similarity score above a given threshold were stored with their corresponding target sentences. In a subsequent study, Jitao et al. (2020) improved the previously proposed method by explicitly informing models about any relevant tokens in the fuzzy matches and incorporating distributed sentence representations (see Table 1).

Inspired by this previous work, we investigate the use of Iconclass in the context of artwork title translations. We use the definitions and keywords associated with Iconclass codes to extend and augment the artwork titles. Our main contribution is that we demonstrate that Iconclass definitions, when provided within a data augmentation strategy, improve translation performance.

extended context	source	cc_sieh cc_, cc_Bob cc_! -Wo sind sie?
	target	-Where are they?
data augmentation	source	How long does a cold last? || Combien de temps dure le vol?
	target	Combien de temps dure un rhume?

Table 1: Examples of the *extended context* approach from Tiedemann and Scherrer (2017) and the *data augmentation* approach from Jitao et al. (2020). The special prefix cc_ indicates tokens from the extended context and the special token || separates the augmented sentence from the main and additional parts.

3 Methods

In this section, we describe the datasets and the methods that we utilized in our research. We justify our choice for the particular NMT model used and provide details on the experimental settings and evaluation measures.

3.1 Datasets and Preprocessing

3.1.1 Iconclass

Iconclass is an iconographic classification system used by stakeholders in the GLAM sector (Galleries, Libraries, Archives and Museums) to describe and retrieve subjects represented in the visual arts. Each subject represented in Iconclass is assigned a unique Iconclass code or identifier, that includes keywords and definitions in multiple languages (see Figure 1). In total, Iconclass contains a set of 28,000 hierarchically ordered definitions and 14,000 keywords. An Iconclass code starts with a digits ranging from 0 to 9 representing 10 main categories: (0) abstract art; (1-5) general topics; (6) history; (7) Bible; (8) literature; (9) classical mythology and ancient history. An Iconclass code can be further complemented by the options presented in Table 2. Keywords have been added to concepts to help users to retrieve relevant concepts from the database. However, the number of available keywords per language differs significantly and they do not necessary match each other across languages, as is evident from the example Iconclass codes in Figures 1b and 1c. Unfortunately, Dutch descriptions are mostly unavailable and, hence, are not utilized in this work. We extracted all information from the Iconclass codes using the Iconclass Python package[1].

N	**Extension**	**Definition**	**Keywords**
1	71H7	David and Bathsheba (2 Samuel 11-12)	Bathsheba, Samuel-2 11-12
	71H7**1**	David, from the roof (or balcony) of his palace, sees Bathsheba bathing	balcony, bathing, love at first sight, palace, roof, spying
2	25G41	flowers	flower
	25G41**(ROSE)**	flowers: rose	rose
3	25F23(LION)	beasts of prey, predatory animals: lion	lion
	25F23(LION)**(+12)**	beasts of prey, predatory animals: lion (+ heraldic animals)	Wappentier, araldica, heraldisches Symbol, heraldry, héraldique, lion

Table 2: Extension of Iconclass codes: (1) a letter or digit to increase specificity; (2) bracketed text to add the name of a specific entity; (3) bracketed text with a plus-sign to add an additional 'shade of meaning'.

3.1.2 Artwork dataset

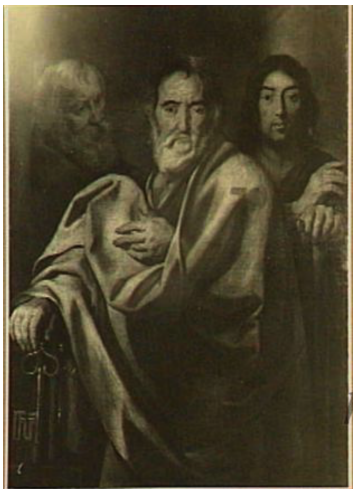
The NL↔EN artwork dataset used below has been extracted from the database of the Netherlands Institute for Art History[2]. We deleted all duplicates from the dataset and finally obtained 21,988 sentence pairs, with the corresponding Iconclass codes for the artworks in question. We randomly selected 2,000 sentence pairs as a development set and included another 2,000 sentence pairs in the test set. The training set contains 1.35 ± 0.72 Iconclass codes per sentence/title and we randomly sample one Iconclass code per sentence/title in the development and test sets. In this work, we do not exploit the hierarchical structure of Iconclass codes and leave this worthwhile option to future work. Additional details about the datasets are provided in Table 3. We experimented with 4 different concatenation strategies. Each source sentence s_i (in English or Dutch) was concatenated using bracketed tags to the corresponding English description d_i^{en} or set of keywords in English k_i^{en} or Dutch k_i^{nl}: (1) s_i (txs) d_i^{en}

[1] https://labs.brill.com/ictestset/

[2] https://rkd.nl/en/explore/images

(a) 11F25

(b) 31A235



(c) 11I35

Figure 1: Examples of images assigned Iconclass codes from Posthumus (2020). The Iconclass code 11F25 provides the English definition 'Mater Dolorosa', the set of English keywords 'Mater Dolorosa, bust, full length, half-length, head , mother' and the set of the Dutch keywords 'Mater Dolorosa, buste, full length , half-length, hoofd, moeder'. The Iconclass code 31A235 is less informative with the definition 'sitting figure' and only one English keyword 'sitting'. The Iconclass code 11I35 only has an English definition: 'other groups of apostles'.

Split	EN sentence	NL sentence	EN description	EN keywords	NL keywords
Train	52.01 ± 29.64	53.75 ± 29.86	51.80 ± 34.60	26.58 ± 21.48	26.85 ± 21.78
Dev	51.90 ± 29.90	53.81 ± 30.37	52.20 ± 34.75	26.42 ± 21.65	26.84 ± 21.65
Test	51.76 ± 29.67	52.61 ± 35.05	52.61 ± 35.05	26.66 ± 21.10	26.95 ± 21.41

Table 3: Statistics of the dataset: mean and standard deviation in sentence lengths.

(txe); (2) s_i (kws) k_i^{en} (kwe); (3) s_i (kws) k_i^{nl} (kwe); (4) s_i (txs) d_i^{en} (txe) (kws) k_i^{en} (kwe). If a sentence from the training set has more than one Iconclass code, then the sentence is separately matched with each Iconclass code using one of the concatenations as shown in Table 5. We balance each concatenated sentence by adding its original version to the training and development sets in order to make the models learn both types of sentences equally well. Additionally, we use two versions of the test set in our evaluation: (1) a baseline test set, where we use original sentences without any concatenations; (2) the test set as concatenated with the corresponding additional information.

3.1.3 Pre-training

Pre-training has become a common solution to cope with small datasets across many domains in deep learning (Ruder et al., 2019). In our experiments, we used 1,777,653 sentence pairs extracted from the Europarl corpus (Tiedemann, 2012) for the NL↔EN language pair, in order to pre-train the models on a generic background corpus. We randomly selected 3,000 sentence pairs as a development set and 3,000 sentence pairs in the test set respectively.

3.2 Model Details

Banar et al. (2020) demonstrated the advantages of character-level translation over the subword-level approach for artwork titles. Hence, we exclusively resorted to character-level models in the present work. However, Banar et al. (2020) used a fusion of recurrent and convolutional models (Lee et al., 2017) that has become outdated. The recent emergence of NMT models started with recurrent neural

source	Ceres bespot door Stellio (Metamorphosen 5: 446-461)
target	Mocking of Ceres by Stellio (Metamorphoses 5: 446-461)
English definition	a little boy (Abas, Ascalabus, or Stellio) laughs at Ceres, because she drinks too avidly while she is resting at an old woman's house
English keywords	Ascalabus, boy, drinking, laughing, old woman, thirst
Dutch keywords	Ascalabus, dorst, drinken, jongen, lachend, old woman
concatenation 1	Ceres bespot door Stellio (Metamorphosen 5: 446-461) (txs) a little boy (Abas, Ascalabus, or Stellio) laughs at Ceres, because she drinks too avidly while she is resting at an old woman's house (txe)
concatenation 2	Ceres bespot door Stellio (Metamorphosen 5: 446-461) (kws) Ascalabus, boy, drinking, laughing, old woman, thirst (kwe)
concatenation 3	Ceres bespot door Stellio (Metamorphosen 5: 446-461) (kws) Ascalabus, dorst, drinken, jongen, lachend, old woman (kwe)
concatenation 4	Ceres bespot door Stellio (Metamorphosen 5: 446-461) (txs) a little boy (Abas, Ascalabus, or Stellio) laughs at Ceres, because she drinks too avidly while she is resting at an old woman's house (txe)(kws) Ascalabus, boy, drinking, laughing, old woman, thirst (kwe)

Table 4: Example concatenations of a title and the information from the Iconclass code 92M132.

source	Johannes de Doper en de H. Hieronymus
target	John the Baptist and St. Jerome
definition 1	the monk and hermit Jerome (Hieronymus); possible attributes: book, cardinal's hat, crucifix, hour-glass, lion, skull, stone
definition 2	John the Baptist; possible attributes: book, reed cross, baptismal cup, honeycomb, lamb, staff
training sentence 1	Johannes de Doper en de H. Hieronymus (txs) the monk and hermit Jerome (Hieronymus); possible attributes: book, cardinal's hat, crucifix, hour-glass, lion, skull, stone (txe)
training sentence 2	Johannes de Doper en de H. Hieronymus (txs) John the Baptist; possible attributes: book, reed cross, baptismal cup, honeycomb, lamb, staff (txe)

Table 5: Example of the concatenation of a title having multiple Iconclass codes with corresponding English definitions from 11H(JEROME) and 11H(JOHN THE BAPTIST).

networks such as GRU or LSTM memory cells (Bahdanau et al., 2014; Sutskever et al., 2014; Luong et al., 2015; Cho et al., 2014), but since then it has been established that Transformer-based architectures (Vaswani et al., 2017) persuasively outperform recurrent and convolutional models across various tasks. The Transformer model mitigates some of the limitations of recurrent and convolutional models; the Transformer, for example, includes self-attention mechanisms that can access all positions in a previous layer. Therefore, the receptive field is not as myopic as with convolutional models. Additionally, the absence of recurrent connections allows one to make the training process fully parallellizable. Therefore, such models nowadays are more appealing for our problem.

3.3 Training and Inference Details

As character-level translation works better for the translation of artwork titles (Banar et al., 2020), we applied a four-layer character-level Transformer (Vaswani et al., 2017), implemented in the OpenNMT-py framework (Klein et al., 2017). The vocabulary size was set to 300 characters and the length of sentences was limited to 450 characters. The models were trained by minimizing the negative conditional

log-likelihood using the Adam optimizer (Kingma and Ba, 2014) with a batch size of 6,144 tokens and an accumulation count of 4. First, we pre-train the models on the general corpus and, then, fine-tune them on our domain-specific corpus. Each model was trained on a single GeForce GTX 1080 Ti with 11 GB RAM. In the pre-training phase, the models were initialized using the method proposed by Glorot and Bengio (2010) and trained for 100,000 updates using the Noam decay schedule (Popel and Bojar, 2018) with an initial learning rate of 2. In the fine-tuning phase, the initial learning rate was set to 0.0001. The fine-tuning was interrupted as soon as the validation loss did not decrease for 600 updates. In the decoding part of the architecture, we applied a beam search with a beam size of 25. The evaluation was conducted using three standard metrics: CHARACTER[3] (Wang et al., 2016), CHRF[4] (Popović, 2015) and BLEU-4 (Papineni et al., 2002) .

4 Results and Discussion

We present our quantitative results in Section 4.1. We divide our experimental results into two different sections. First, we assess the use of the extended context in Section 4.1.1. Second, we compare the various data augmentation strategies to the baseline in Section 4.1.2. In Section 4.2 we manually inspect a selection of outputs for the best performing model.

4.1 Quantitative Analysis

Pair	Additional Information	Type	Baseline test			Test with context		
			BLEU↑	C-TER↓	CHRF↑	BLEU↑	C-TER↓	CHRF↑
NL-EN	(a) baseline	NA	**43.25**	30.71	67.04	NA	NA	NA
	(b) keywords^{nl}	EC	42.50	30.81	66.77	42.56	30.60	66.91
	(c) keywords^{en}	DA	42.93	**30.25**	**67.21**	42.96	30.03	67.31
	(d) definition^{en}	DA	42.25	30.58	66.77	**46.02**	**29.48**	**68.32**
	(e) (d) + (c)	DA	41.96	31.23	66.47	45.88	29.79	68.17
EN-NL	(f) baseline	NA	42.99	30.82	67.75	NA	NA	NA
	(g) keywords^{nl}	DA	43.83	30.37	68.21	**43.71**	30.22	68.18
	(h) keywords^{en}	EC	43.59	30.49	68.07	43.53	30.66	67.92
	(i) definition^{en}	EC	42.93	30.94	67.73	43.20	30.80	67.77
	(j) (i) + (h)	EC	**44.00**	**30.35**	**68.23**	43.68	**29.97**	**68.37**

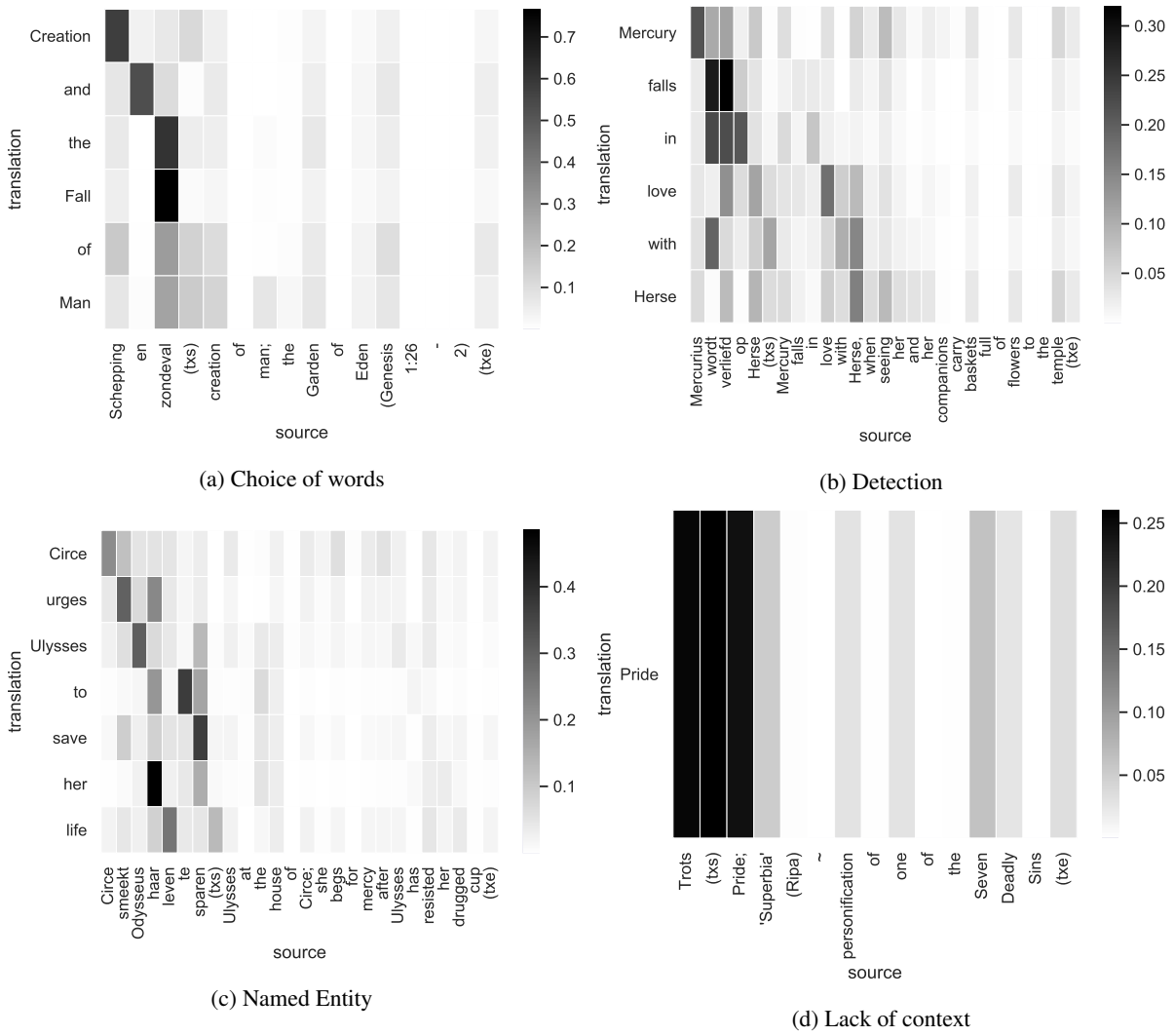
Table 6: Results of the experiments. The type 'EC' and 'DA' correspond to experiments with *extended context* and *data augmentation* approaches respectively. The arrows near the metrics in the column labels indicate the desired direction of improvement (i.e. whether a higher/lower score for this metric is better). The 'baseline' experiment corresponds to the models fine-tuned without any additional information. The columns 'Test with context' and 'Baseline test' correspond to translation of the sentences enriched with additional information and without it, correspondingly.

4.1.1 Extended Context

From Table 6 we can see that the models (b, h) fine-tuned with the context extended by keywords demonstrate comparable results to the baseline models (a, f) in the baseline test. Model (h) slightly outperforms the associated baseline, while model (b) is slightly worse in this testing scenario. The testing scenario with the extended context suggests that these models do not successfully manage to exploit the additional context, as there is no obvious boost in performance. For model (h), we can even observe a subtle decrease in performance. As the definitions were only available in English, the context has been extended by definitions only for EN→NL (see the models i and j). These models still show comparable performance to the baseline model (f) when translating without the extended context. The testing scenario with the extended context is also not beneficial. Therefore, we observe that the extended

[3] https://github.com/rwth-i6/CharacTER

[4] https://github.com/m-popovic/chrF



(a) Choice of words

(b) Detection

(c) Named Entity

(d) Lack of context

Figure 2: Example translations with attention maps for the model that uses English Iconclass definitions for data augmentation. We also provide the target sentence, translation of the model without data augmentation (**Without DA**) and translation of the baseline model (**Baseline**) fine-tuned without additional information. (a) **Target:** 'Creation and the fall of Man'. **Without DA:** 'Creation and a fall of mankind'. **Baseline:** 'Creation and a fall of mankind'. (b) **Target:** 'Mercury falls in love with Herse'. **Without DA:** 'Mercury being loved on Herse' **Baseline:** 'Mercury being loved on Herse'. (c) **Target:** 'Circe begging Ulysses to save her life'. **Without DA:** 'Circe pleading with Odysseus to save her life'. **Baseline:** 'Circe pleads with Odysseus to save her life in her life'. (d) **Target:** 'Pride'. **Without DA:** 'Christ'. **Baseline:** 'Events'.

context in the testing phase is generally not advantageous and, hence, we conclude that the use of these concatenations is not beneficial in our case.

4.1.2 Data Augmentation

As shown in Table 6, the models (c, g) fine-tuned with keywords as a source of data augmentation show comparable results to the baseline models (a, f) in the baseline testing scenario. Therefore, we conclude that the source sentences augmented by keywords are not helpful for our task. The models (d, e) utilize definitions for the data augmentation for NL→EN. These models demonstrate slightly worse performance compared to the baseline model (a) when translating without data augmentation. Translating with the source sentences augmented by definitions, however, boosts the performance

of these models and we observe substantial improvement over baseline model (a) and their own non-augmented counterparts. We manually inspected the output of the model (d) and discuss some interpretive observations in the next Section 4.2.

4.2 Error Analysis

In this section, we qualitatively discuss the outputs of the best performing model (d) in the testing scenario with data augmentation from Table 6. We compare its output to that for the basic test setting, as well as to baseline model (a). We divide our findings into four main categories below.

Lexical choices. The model is able to exploit additional information in the target language in order to select words closer to target translations. From Figure 2a, we can see that the target translation is not literal and the word *Man* is absent in the Dutch counterpart. The literal translation of this sentence is *Creation and Fall*. In this case, baseline model (a) and model (d) in the basic testing scenario generate non-literal translations with the additional post-modification *of mankind*. The translation is close to the target, but it is still erroneous. Model (d) pays attention to the phrase *creation of man* from the Iconclass definition and decides to adopt the word *Man* instead of the word *mankind* as in the baseline and in the case without data augmentation.

Detection of lexical units. As mentioned in Section 3.1.1, Iconclass definitions describe subjects represented in artwork images. If a subject is widely represented in iconography, an artwork can even have the same title as an Iconclass definition or the definition can at least contain parts of the target translations. Hence, these target translations may be detected in an Iconclass definition and just copied by model. In Figure 2b, we can see that the baseline model (a) and model (d) in the basic testing scenario produce grammatically incorrect outputs. The Dutch fixed expression *wordt verliefd op* is translated almost literally as *being loved on*. However, we can see that model (d) finds a part of the right translation in the additional information and copies it.

Named entities. Artwork titles densely feature named entities in comparison to general corpora and, hence, they can be a serious issue for NMT models. Similarly to Banar et al. (2020), we observe that the models in the basic testing scenario tend to copy named entities instead of attempting a proper translation. However, we observe that if a correct named entity is provided in the additional information, the model is able to generate the correct target translation. From Figure 2c, we can see that model (d) derives the correct named entity *Ulysses*, while other scenarios are less successful.

Lack of context. The titles of artworks naturally differ from the sentences in more general corpora and can be very short. The lack of context may cause translation difficulties. In the example from Figure 2d, the title consists of only one word. The baseline model (a) and model (d) in the basic testing scenario struggle to translate the title *Pride* correctly and generate the non-sense translations *Events* and *Christ*, correspondingly. In this case, the additional information helps model (d) to generate the right answer.

5 Conclusion and Future Work

In this paper, we utilized the Iconclass framework as a source of additional information for NMT of artwork titles. We extracted English and Dutch keywords and English definitions from the Iconclass codes. This information was concatenated to the source sentences. Experiments show that augmenting the source sentences with the Iconclass definitions for the objects under scrutiny improves the overall translation quality by a considerable margin. On the basis of a manual inspection of the output, we argue that the NMT model is able to successfully capitalize on the additional information extracted from the Iconclass definitions. There are various reasons for this. Firstly, the model is able to recognize any named entities in the concatenated part (that are lacking in the actual source title) and it can correctly inject them in the translation. Secondly, the data augmentation approach improves the lexical aspects of the translation, providing useful semantic cues in the case of limited context. Thirdly, the model is able to detect correct translations in the concatenated part and integrate them appropriately in the translation. However, the augmentation of the source sentences with the keywords and any type of extended context that we applied do not show any promising results. In future work, we would like to extend the current

pipeline with a model that automatically matches the source sentence with the corresponding Iconclass code. In addition, it may be beneficial to incorporate visual features extracted from artworks and, hence, to perform a multi-modal matching. Iconclass contains 28,000 hierarchically ordered definitions that makes the matching problem extremely sophisticated. We plan to investigate the feasibility of such a matching strategy.

References

Ruchit Rajeshkumar Agrawal, Marco Turchi, and Matteo Negri. 2018. Contextual handling in neural machine translation: Look behind, ahead and on both sides. In *21st Annual Conference of the European Association for Machine Translation*, pages 11–20.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Nikolay Banar, Karine Lasaracina, Walter Daelemans, and Mike Kestemont. 2020. Transfer learning for digital heritage collections: Comparing neural machine translation at the subword-level and character-level. In *Proceedings of the 12th International Conference on Agents and Artificial Intelligence - Volume 1: ARTIDIGH,*, pages 522–529. INSTICC, SciTePress.

Rachel Bawden, Rico Sennrich, Alexandra Birch, and Barry Haddow. 2018. Evaluating discourse phenomena in neural machine translation. In *16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1304–1313.

Bram Bulté and Arda Tezcan. 2019. Neural fuzzy repair: Integrating fuzzy matches into neural machine translation. In *57th Conference of the Association for Computational Linguistics (ACL)*, pages 1800–1809.

Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder–decoder approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 103–111.

Marco Fiorucci, Marina Khoroshiltseva, Massimiliano Pontil, Arianna Traviglia, Alessio Del Bue, and Stuart James. 2020. Machine learning for cultural heritage: A survey. *Pattern Recognition Letters*, 133:102–108.

Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256.

XU Jitao, Josep M Crego, and Jean Senellart. 2020. Boosting neural machine translation with similar translations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1580–1590.

Marcin Junczys-Dowmunt. 2019. Microsoft translator at wmt 2019: Towards large-scale document-level neural machine translation. In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 225–233.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M. Rush. 2017. OpenNMT: Open-source toolkit for neural machine translation. In *Proc. ACL*.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature*, 521(7553):436–444.

Jason Lee, Kyunghyun Cho, and Thomas Hofmann. 2017. Fully character-level neural machine translation without explicit segmentation. *Transactions of the Association for Computational Linguistics*, 5:365–378.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer.

Minh-Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics.

Martin Popel and Ondřej Bojar. 2018. Training tips for the transformer model. *The Prague Bulletin of Mathematical Linguistics*, 110(1):43–70.

Maja Popović. 2015. chrf: character n-gram f-score for automatic mt evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395.

Etienne Posthumus. 2020. Brill iconclass ai test set. `https://labs.brill.com/ictestset/`.

Sebastian Ruder, Matthew E Peters, Swabha Swayamdipta, and Thomas Wolf. 2019. Transfer learning in natural language processing. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials*, pages 15–18.

Jürgen Schmidhuber. 2015. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

Jörg Tiedemann and Yves Scherrer. 2017. Neural machine translation with extended context. In *Proceedings of the Third Workshop on Discourse in Machine Translation*, pages 82–92.

Jörg Tiedemann. 2012. Parallel data, tools and interfaces in opus. In *Lrec*, volume 2012, pages 2214–2218.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

G. Vellekoop, E. Tholen, and L. D. Couprie. 1973. *Iconclass : an iconographic classification system*. North-Holland Pub. Co., Amsterdam.

Weiyue Wang, Jan-Thorsten Peter, Hendrik Rosendahl, and Hermann Ney. 2016. Character: Translation edit rate on character level. In *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*, pages 505–510.