


<p>An Artist Making a Powerful Statement — by Creating Work About Herself</p>  <p>During the final days of her solo show at Kravets Wehby Gallery in Manhattan this past spring, the mixed-media artist Theresa Chromati had something to confess about her latest body of work. “I realized that you can’t hide from anything,” she said, staring up at the 2019 painting “We All Look Back at It (Morning Ride).”</p> <p>...</p>	<table border="1"> <tr> <td>Ground-truth caption</td><td>The mixed-media artist Theresa Chromati sits in front of an unfinished and currently untitled acrylic painting at her Brooklyn studio.</td></tr> <tr> <td>LSTM + weighted RoBERTa</td><td>“The Last Man,” 2018, by Theresa Brandonati.</td></tr> <tr> <td>LSTM + GloVe</td><td>“The Red Book,” a work by the artist and artist J.W. Anderson.</td></tr> <tr> <td>Transformer + GloVe</td><td>Nina Hoss in her solo show “Sleeping Beauty” at the Gagosian Gallery.</td></tr> <tr> <td>Transformer + weighted RoBERTa</td><td>Theresa Cromati, “Untitled (The New York Times)” (2016).</td></tr> <tr> <td>+ context</td><td>Theresa Cromati, who has created a new work, “We All Look Back at It (Morning Ride),” 2019.</td></tr> <tr> <td>+ face attention</td><td>Theresa Chromati at the Kravets Wehby Gallery in Manhattan.</td></tr> <tr> <td>+ copying</td><td>Theresa <u>Chromati</u>, who has a new work, “We All Look Back at It (Morning Ride),” 2019.</td></tr> </table>	Ground-truth caption	The mixed-media artist Theresa Chromati sits in front of an unfinished and currently untitled acrylic painting at her Brooklyn studio.	LSTM + weighted RoBERTa	“The Last Man,” 2018, by Theresa Brandonati .	LSTM + GloVe	“The Red Book,” a work by the artist and artist J.W. Anderson .	Transformer + GloVe	Nina Hoss in her solo show “Sleeping Beauty” at the Gagosian Gallery .	Transformer + weighted RoBERTa	Theresa Cromati , “Untitled (The New York Times)” (2016).	+ context	Theresa Cromati , who has created a new work, “We All Look Back at It (Morning Ride),” 2019.	+ face attention	Theresa Chromati at the Kravets Wehby Gallery in Manhattan.	+ copying	Theresa <u>Chromati</u> , who has a new work, “We All Look Back at It (Morning Ride),” 2019.
Ground-truth caption	The mixed-media artist Theresa Chromati sits in front of an unfinished and currently untitled acrylic painting at her Brooklyn studio.																
LSTM + weighted RoBERTa	“The Last Man,” 2018, by Theresa Brandonati .																
LSTM + GloVe	“The Red Book,” a work by the artist and artist J.W. Anderson .																
Transformer + GloVe	Nina Hoss in her solo show “Sleeping Beauty” at the Gagosian Gallery .																
Transformer + weighted RoBERTa	Theresa Cromati , “Untitled (The New York Times)” (2016).																
+ context	Theresa Cromati , who has created a new work, “We All Look Back at It (Morning Ride),” 2019.																
+ face attention	Theresa Chromati at the Kravets Wehby Gallery in Manhattan.																
+ copying	Theresa <u>Chromati</u> , who has a new work, “We All Look Back at It (Morning Ride),” 2019.																

Figure 4: An example from the NYTimes800k test set. The name “Chromati” has never appeared in the training data. Words in blue do not appear in the article and are hallucinated by the decoder. Words highlighted in red are spelling mistakes. Underlined words are those that have been copied by the copying mechanism.

Table 2: NYTimes800k training, validation, and test splits

	Training	Validation	Test
Number of articles	434 272	3 052	8 495
Number of images	764 049	7 852	21 977
Start month	Mar 15	May 19	Jun 19
End month	Apr 19	May 19	Aug 19

there will be events and people in the test data that have never been covered by the news before. In particular, out of the 100K proper nouns in the test captions, 4% never appear in any training captions. Half of these also never appear in any training article. Thus splitting by time allows us study how well the model can generate rare names.

5. Experiments

5.1. Training Details

For parameter optimisation we use the adaptive gradient algorithm Adam [19] with the following parameter settings: $\beta_1 = 0.9$, $\beta_2 = 0.98$, $\epsilon = 10^{-6}$. We warm up the learning rate in the first 5% of the training steps to 10^{-4} , and decay it linearly afterwards.

We apply L_2 regularisation to all network weights with a value of 10^{-5} and use the weight decay fix [25] to decouple

the learning rate from the regularisation parameter. We clip the gradient norm at 0.1. We use a batch size of ***** and train all models for 6.6 million steps. This is equivalent to 16 epochs on GoodNews and 9 epochs on NYTimes800k. Training is done with mixed precision to reduce the memory footprint and allow our full model to be trained on a single GPU. This full model takes 5 days to train on one Titan V GPU and has 207 million trainable parameters – see the supplement for the number of trainable parameters in each model variant.

The training pipeline is written in PyTorch [31] using the AllenNLP framework [14]. The RoBERTa model and dynamic convolution code are adapted from fairseq [29].

5.2. Evaluation Metrics

We use BLEU [30], ROUGE [23], METEOR [7], and CIDEr [44] scores as they are standard for evaluating image captions. These are obtained using the COCO caption evaluation toolkit². Note that CIDEr is particularly suited to evaluating news captioning models as it puts more weight that other metrics on uncommon words. In addition, we evaluate the precision and recall on: named entities, personal names, and rare proper names. Named entities are identified in both the ground truth captions and the generated captions using the SpaCy [] natural language process-

²<https://github.com/tylin/coco-caption>