Massachusetts Institute of Technology

CSAIL

Language Style Transfer

Tianxiao Shen

Different Language Styles

King James Bible

In the beginning God created the heaven and the earth.
And God saw the light, that it was good:
and God divided the light from the darkness.

Bible in basic English

At the first God made the heaven and the earth.
And God, looking on the light, saw that it was good:
and God made a division between the light and the dark.

Simplicity, formality, politeness, personal styles…

Language Style Transfer

King James Bible

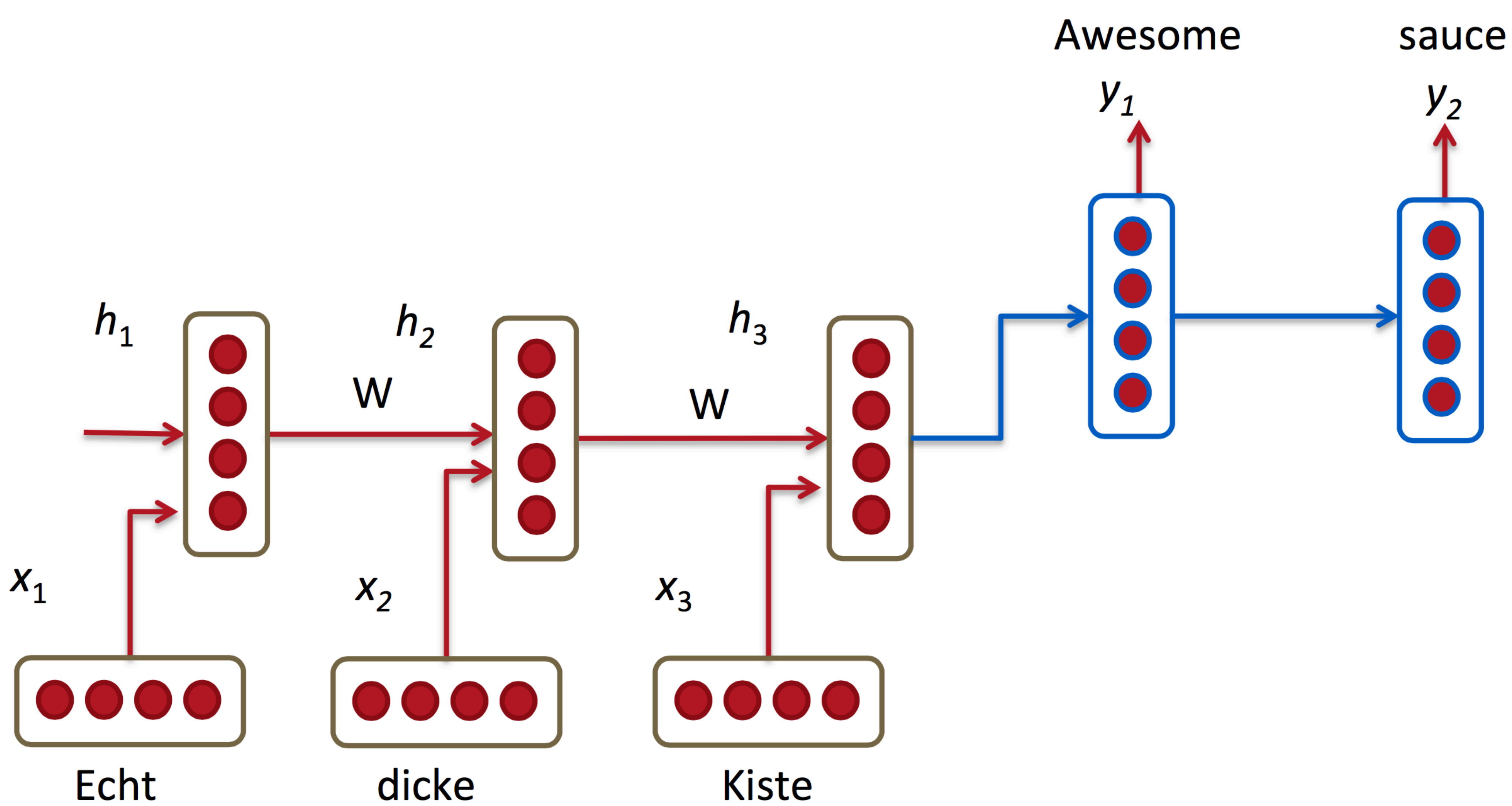
In the beginning God created the heaven and the earth.
And God saw the light, that it was good:
and God divided the light from the darkness.

Bible in basic English

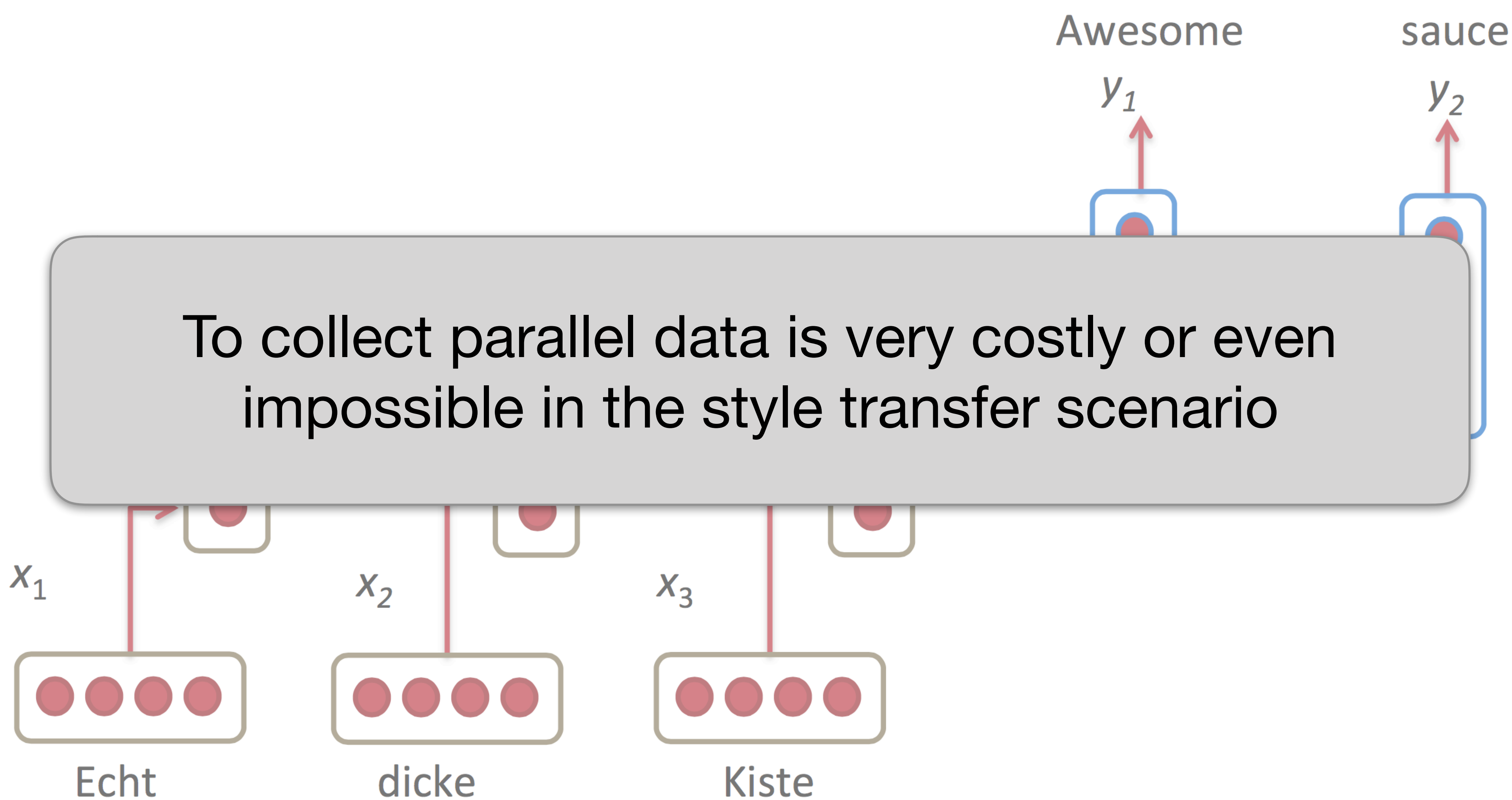
At the first God made the heaven and the earth.
And God, looking on the light, saw that it was good:
and God made a division between the light and the dark.

- Towards real language understanding
- Personalized chatbots, appropriately convey a message according to different social contexts…

Parallel Translation

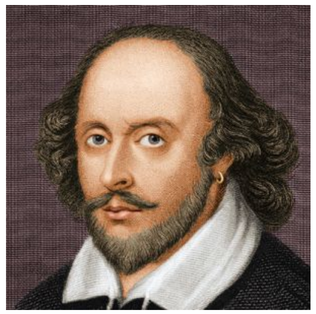


Parallel Translation



To collect parallel data is very costly or even impossible in the style transfer scenario

Non-Parallel Transfer



To be, or not to be, that is the question:
Whether 'tis nobler in the mind to suffer
The slings and arrows of outrageous fortune,
Or to take Arms against a Sea of troubles,
And by opposing end them: to die, to sleep

Donald J. Trump
@realDonaldTrump

They're bringing drugs,
they're bringing crime,
they're rapists, and some,
I assume, are good people

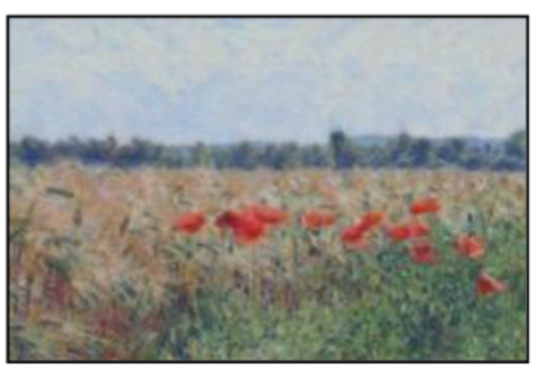
Obama, and all others,
have been so weak,
and so politically correct,
that terror groups are
forming and getting
stronger! Shame.

Image Style Transfer

photograph

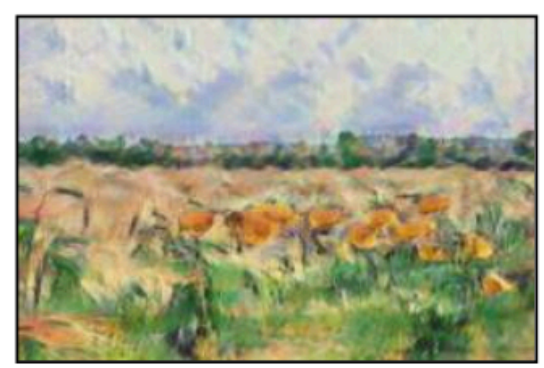
artwork

after transfer

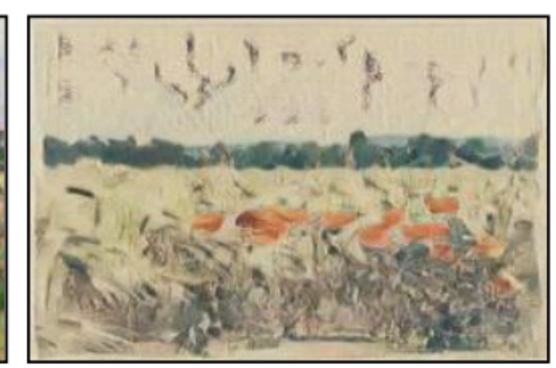


Monet

Van Gogh



Cezanne



Ukiyo-e

[Zhu et al. 2017]

Challenges in Language Style Transfer

- Style and content interact in subtle ways
- Content must be preserved
- Discreteness

Our Approach

- Style and content interact in subtle ways
- Content must be preserved
- Discreteness

- Map between sentences and continuous latent representations
- Decompose latent representations into style and content
- Modify the latent style component to realize style transfer

Generative Assumption

a latent style variable $y \sim p(y)$

a latent content variable $z \sim p(z)$

a sentence $x \sim p(x|y, z)$

We observe two corpora in different styles:

$X_1 = \{x_1^{(1)}, \cdots, x_1^{(n)}\}$ consisting of samples from $p(x|y_1)$

$X_2 = \{x_2^{(1)}, \cdots, x_2^{(m)}\}$ consisting of samples from $p(x|y_2)$

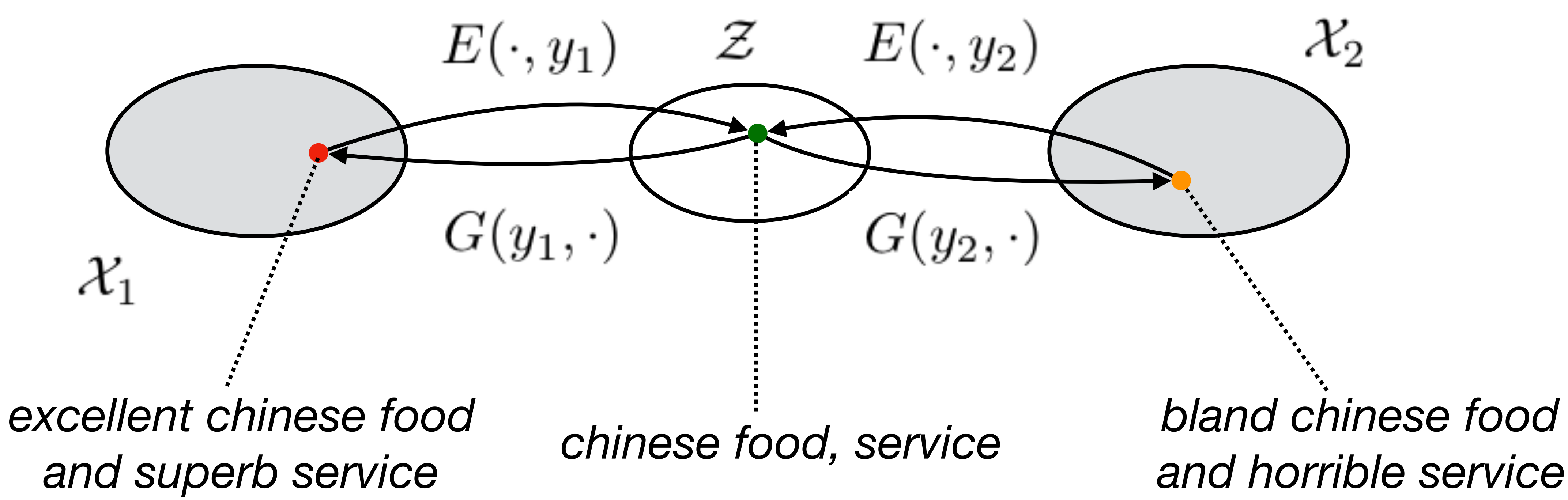
Model Overview

Encoder $E: \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$

to infer the content for a given sentence and style

Generator $G: \mathcal{Y} \times \mathcal{Z} \to \mathcal{X}$

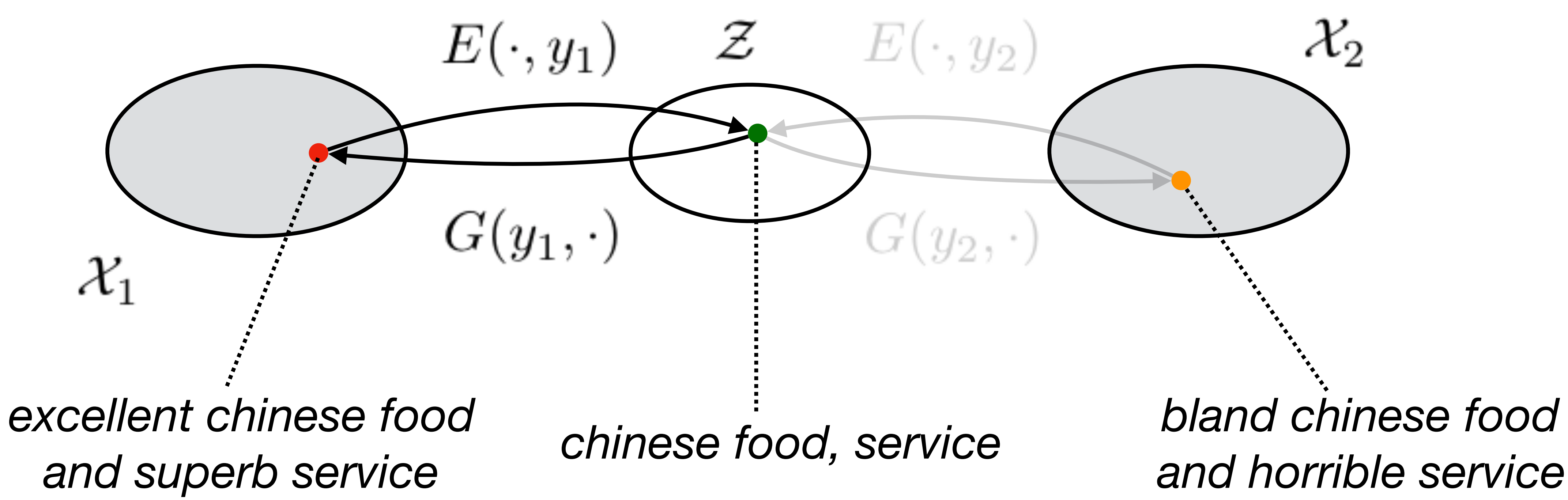
to generate a sentence from a given style and content



Model Overview

E and G form an auto-encoder when applying to the same style

$$G(y_1, \cdot) \circ E(\cdot, y_1) = \mathrm{id}_{\mathcal{X}_1} \quad G(y_2, \cdot) \circ E(\cdot, y_2) = \mathrm{id}_{\mathcal{X}_2}$$



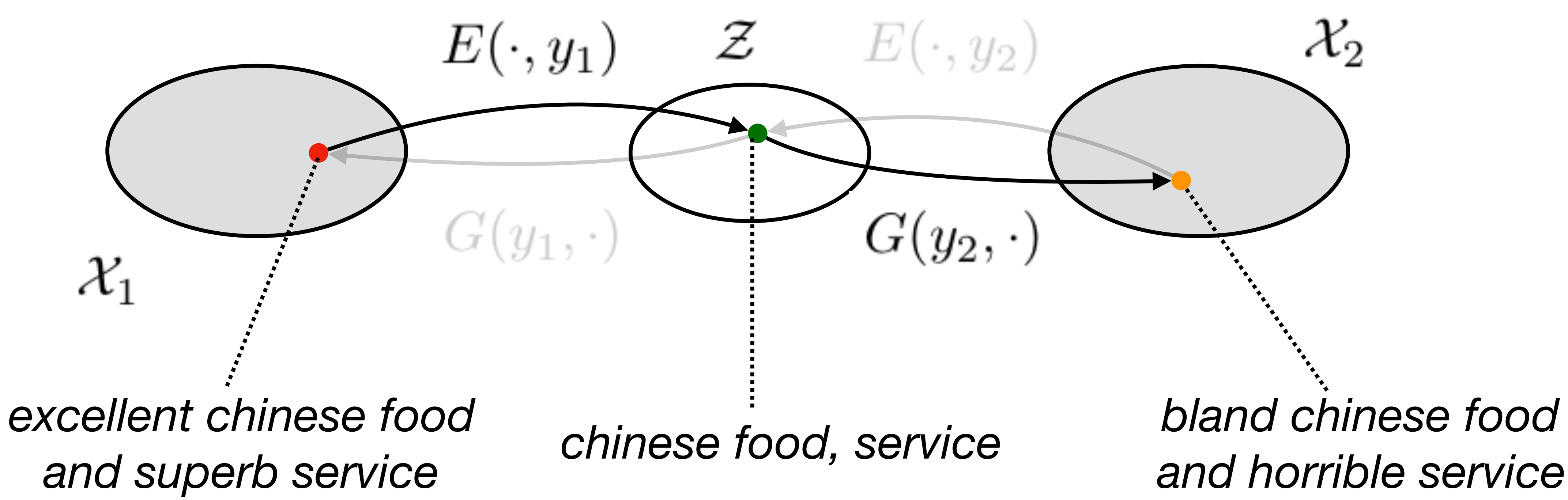
Model Overview

E and G form an auto-encoder when applying to the same style

$$G(y_1, \cdot) \circ E(\cdot, y_1) = \mathrm{id}_{\mathcal{X}_1} \quad G(y_2, \cdot) \circ E(\cdot, y_2) = \mathrm{id}_{\mathcal{X}_2}$$

E and G form a transfer model when applying to different styles

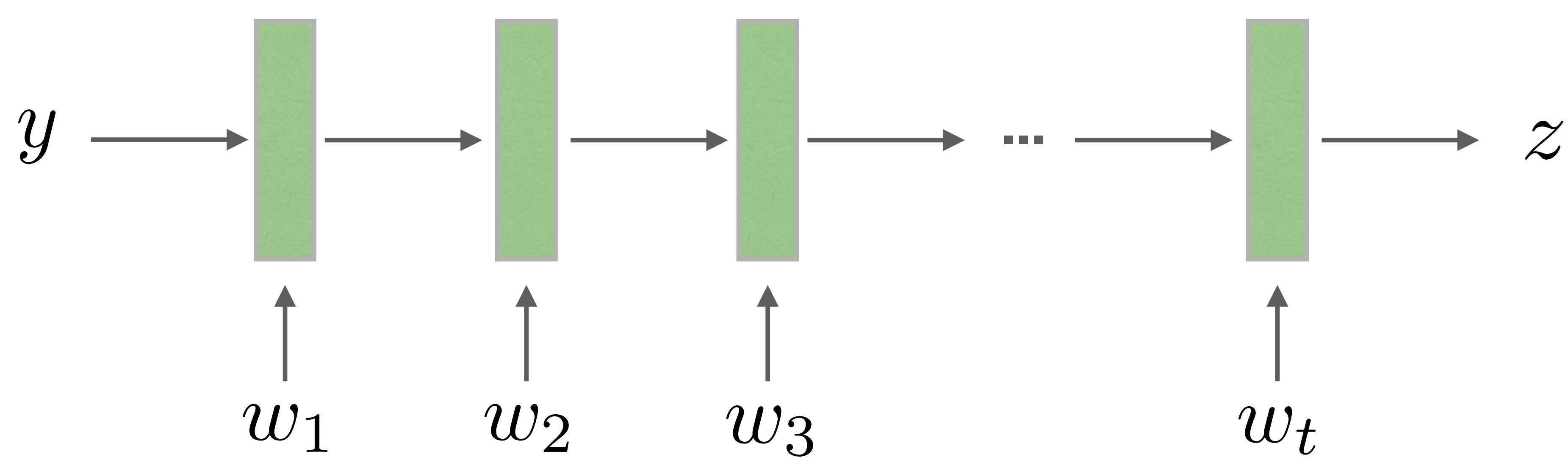
$$G(y_2, \cdot) \circ E(\cdot, y_1) : \mathcal{X}_1 \to \mathcal{X}_2 \quad G(y_1, \cdot) \circ E(\cdot, y_2) : \mathcal{X}_2 \to \mathcal{X}_1$$



Model Architecture

Encoder $E : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$

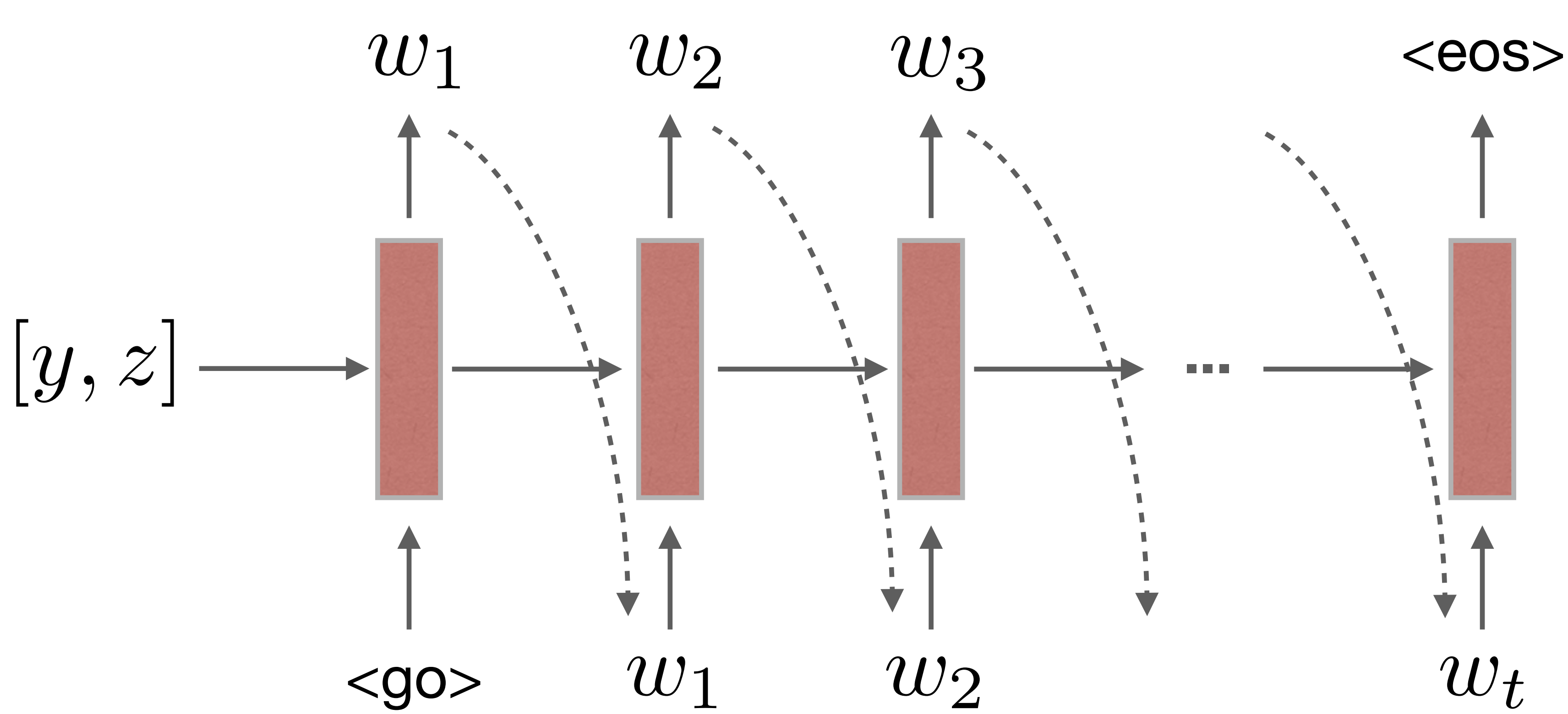
to infer the content for a given sentence and style



Model Architecture

Generator $G : \mathcal{Y} \times \mathcal{Z} \to \mathcal{X}$

to generate a sentence from a given style and content



Reconstruction Loss

E and G form an auto-encoder when applying to the same style

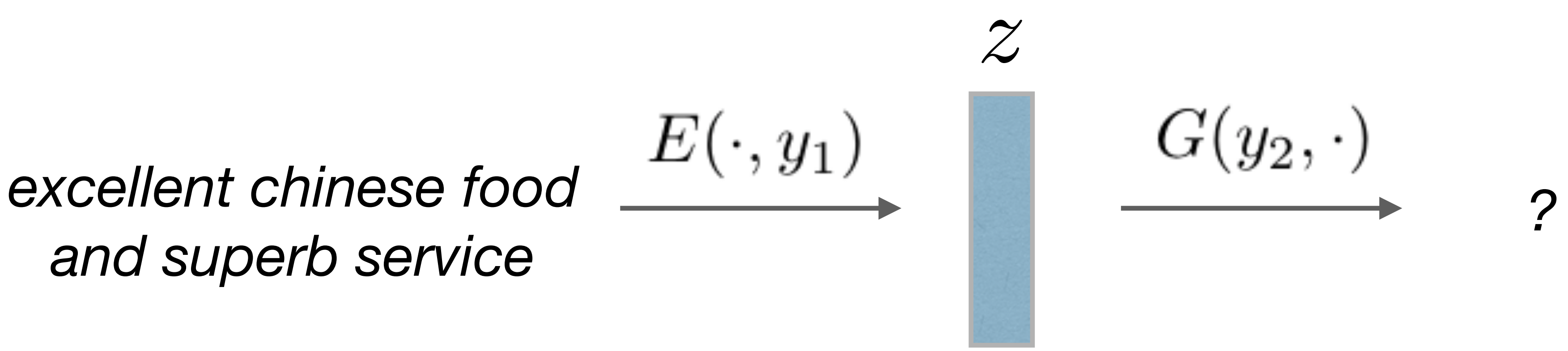
$$G(y_1, \cdot) \circ E(\cdot, y_1) = \mathrm{id}_{\mathcal{X}_1} \quad G(y_2, \cdot) \circ E(\cdot, y_2) = \mathrm{id}_{\mathcal{X}_2}$$

$$\begin{aligned}\mathcal{L}_{\mathrm{rec}}(\theta_E, \theta_G) = \; & \mathbb{E}_{x_1 \sim X_1}[-\log p_G(x_1 | y_1, E(x_1, y_1))] \; + \\ & \mathbb{E}_{x_2 \sim X_2}[-\log p_G(x_2 | y_2, E(x_2, y_2))]\end{aligned}$$

Good to Go?

E and G form a transfer model when applying to different styles

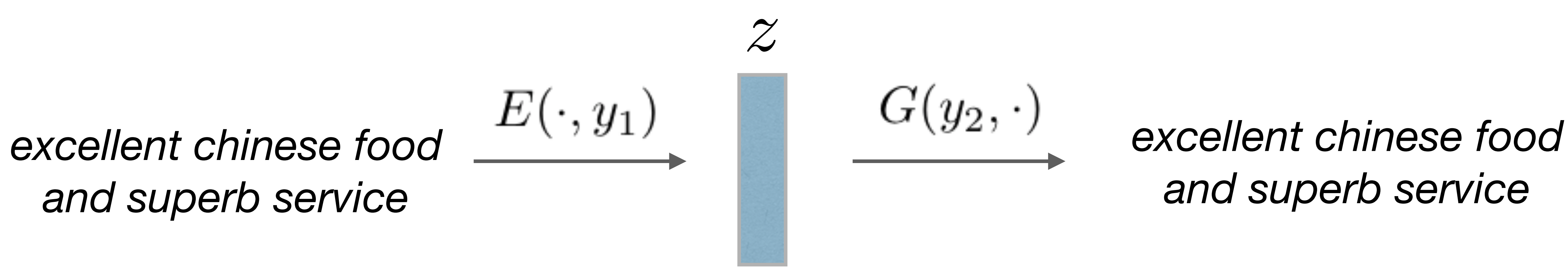
$$G(y_2, \cdot) \circ E(\cdot, y_1) : \mathcal{X}_1 \to \mathcal{X}_2 \qquad G(y_1, \cdot) \circ E(\cdot, y_2) : \mathcal{X}_2 \to \mathcal{X}_1$$



Just Copy, No Transfer

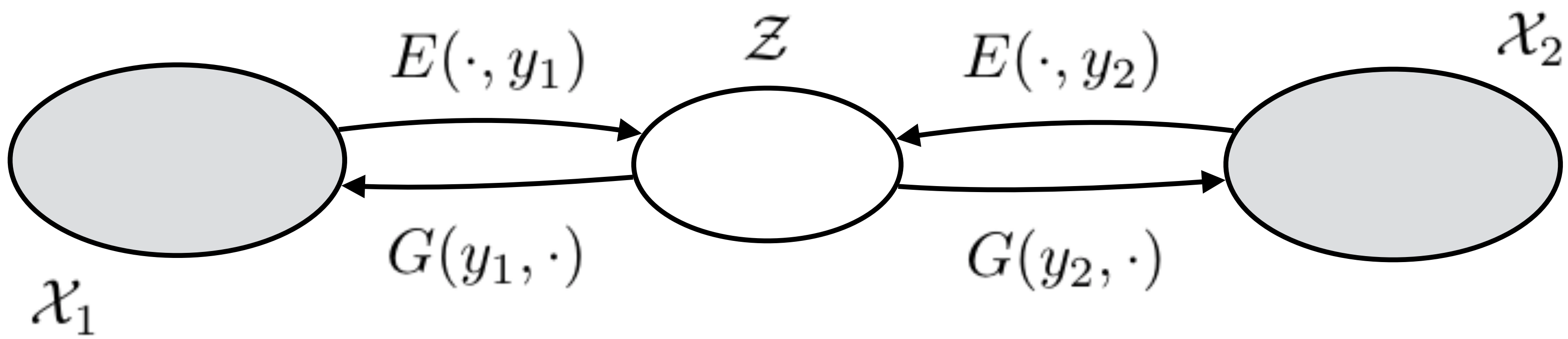
E and G form a transfer model when applying to different styles

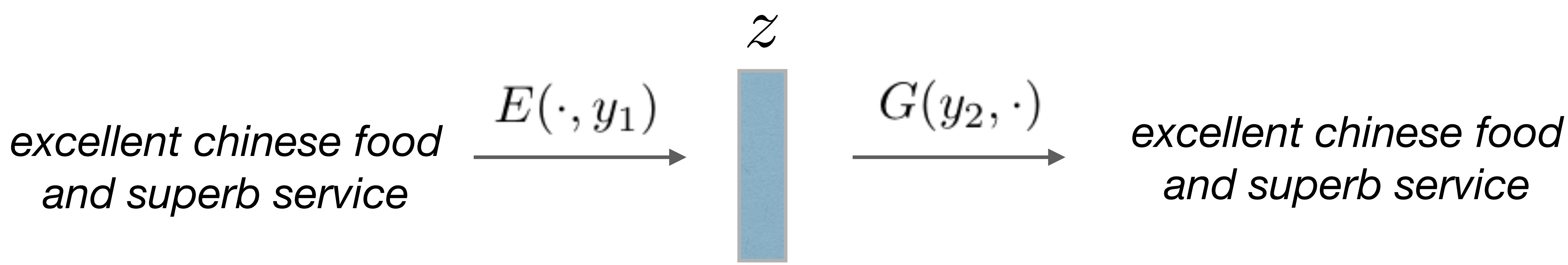
$$G(y_2, \cdot) \circ E(\cdot, y_1) : \mathcal{X}_1 \to \mathcal{X}_2 \quad G(y_1, \cdot) \circ E(\cdot, y_2) : \mathcal{X}_2 \to \mathcal{X}_1$$



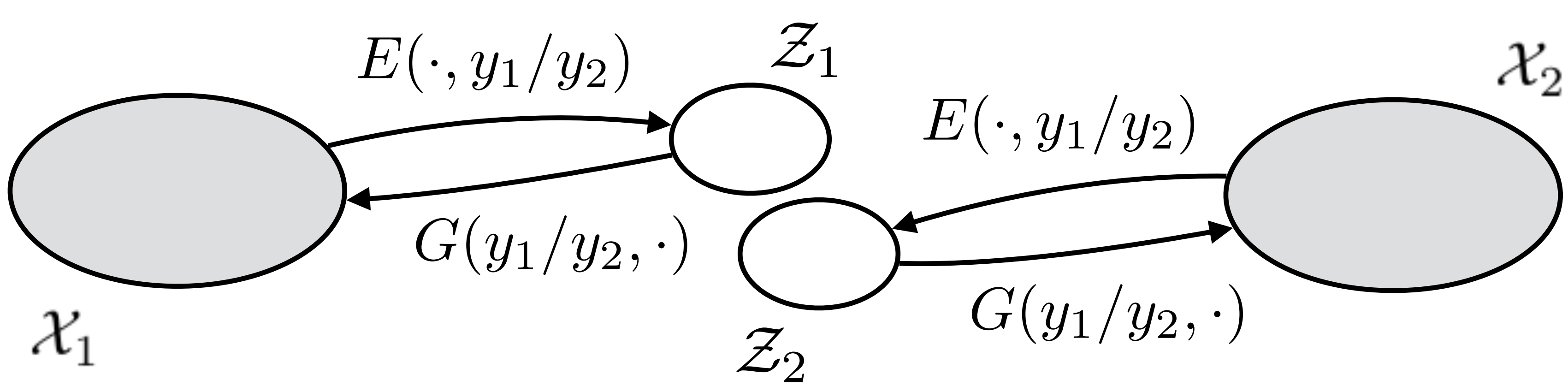
Just Copy, No Transfer

Hope:



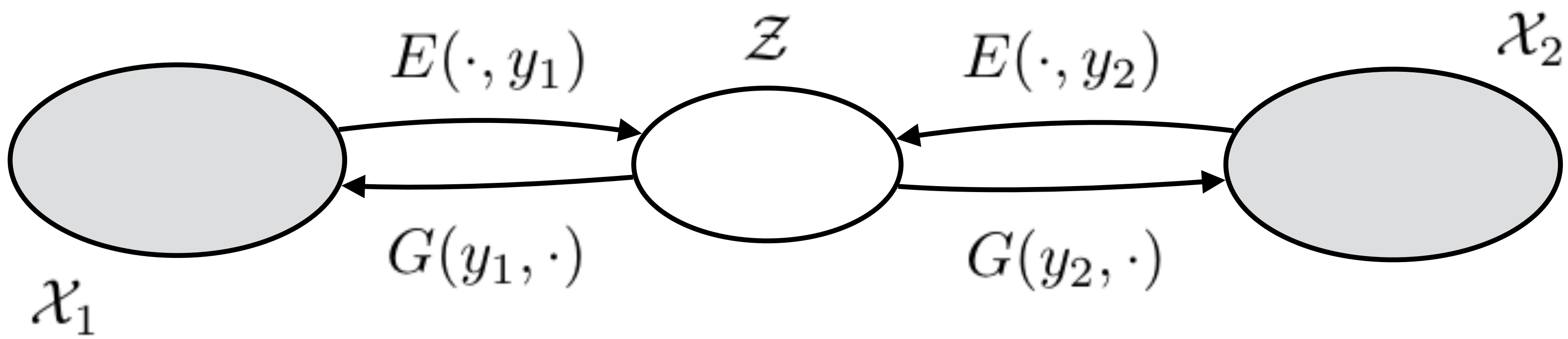


Reality:



0 reconstruction loss

Shared Content Distribution



Constrained optimization problem:

$$\theta^* = \arg\min_{\theta} \mathcal{L}_{\text{rec}}(\theta_E, \theta_G)$$

$$\text{s.t.} \quad E(x_1, y_1) \stackrel{\text{d}}{=} E(x_2, y_2) \qquad x_1 \sim X_1, x_2 \sim X_2$$

Aligned Auto-Encoder

$$\theta^* = \arg\min_{\theta} \mathcal{L}_{\text{rec}}(\theta_E, \theta_G)$$
$$\text{s.t.} \quad E(x_1, y_1) \overset{\text{d}}{=} E(x_2, y_2) \qquad x_1 \sim X_1, x_2 \sim X_2$$

Introduce D to distinguish Z_1 and Z_2:

$$\begin{aligned}\mathcal{L}_{\text{adv}}(\theta_E, \theta_D) = \; & \mathbb{E}_{x_1 \sim X_1}[-\log D(E(x_1, y_1))] + \\ & \mathbb{E}_{x_2 \sim X_2}[-\log(1 - D(E(x_2, y_2)))]\end{aligned}$$

$Z_1 \overset{d}{=} Z_2$ when they're indistinguishable to D

Overall training objective: $\boxed{\min_{E,G} \max_{D} \mathcal{L}_{\text{rec}} - \lambda \mathcal{L}_{\text{adv}}}$

Aligned Auto-Encoder

Results:

great !
horrible !

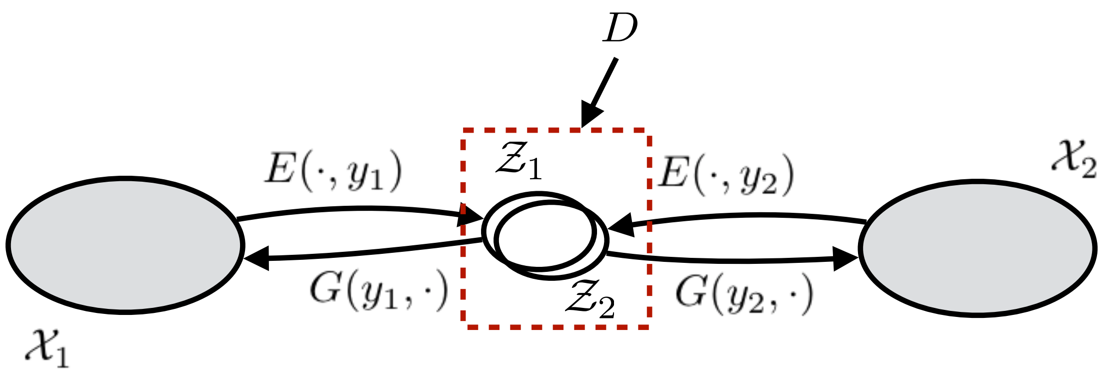
mediocre dim sum if you 're from southern california .
dim sum if you can not choose from california .

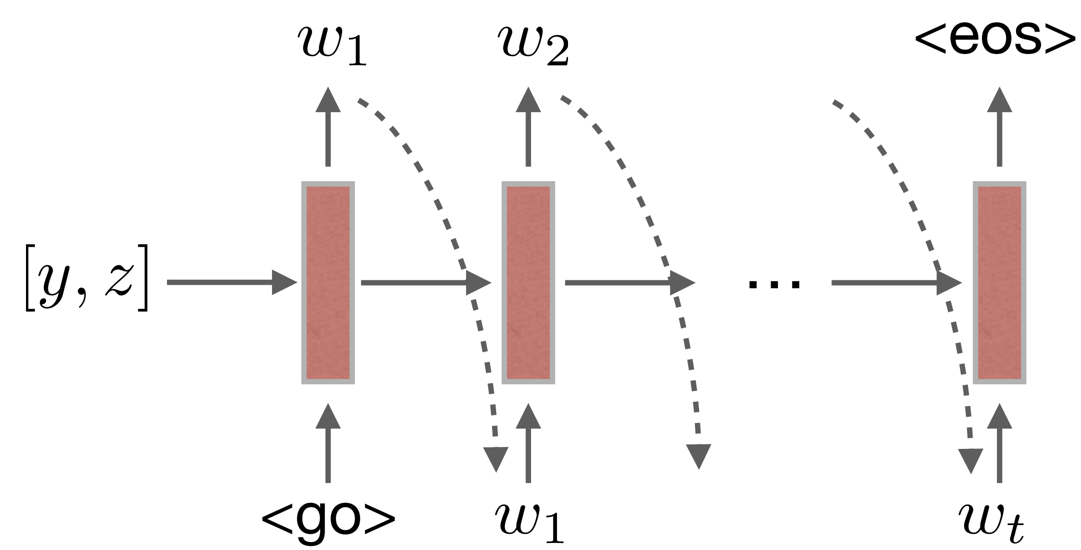
i would n't bother .
i would n't bother .

i would never go back for the food .
i would definitely go back for the food .

- 48.3% sentiment accuracy as measured by a classifier 🤔

Aligned Auto-Encoder

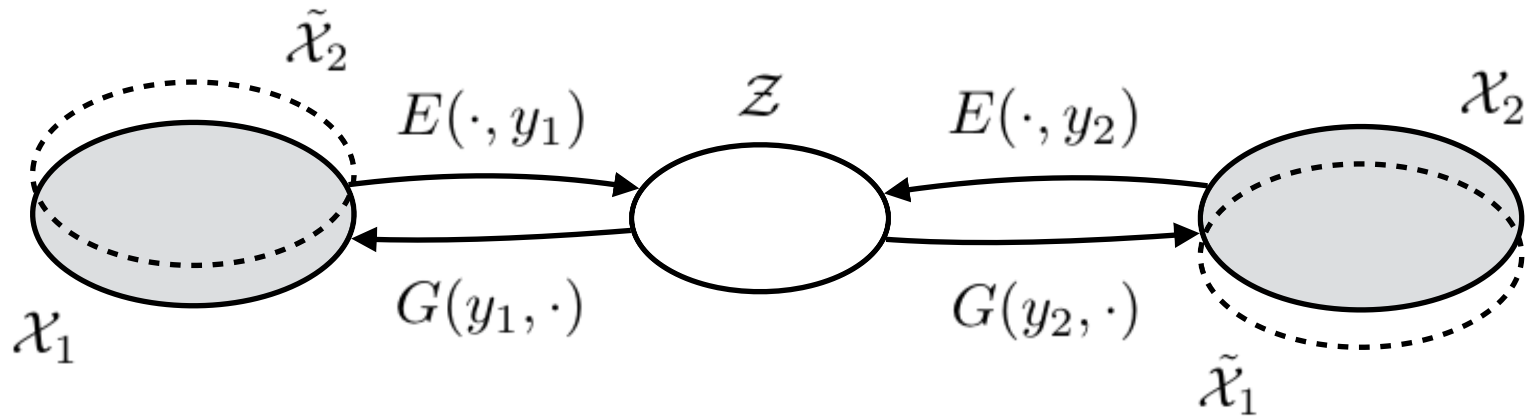




z_1 and z_2's initial misalignment could propagate through the recurrent generating process

As a result the transferred sentence may end up somewhere far from the target domain

Cross Alignment



Transferred sentences from one style should match example sentence from the other style as a population

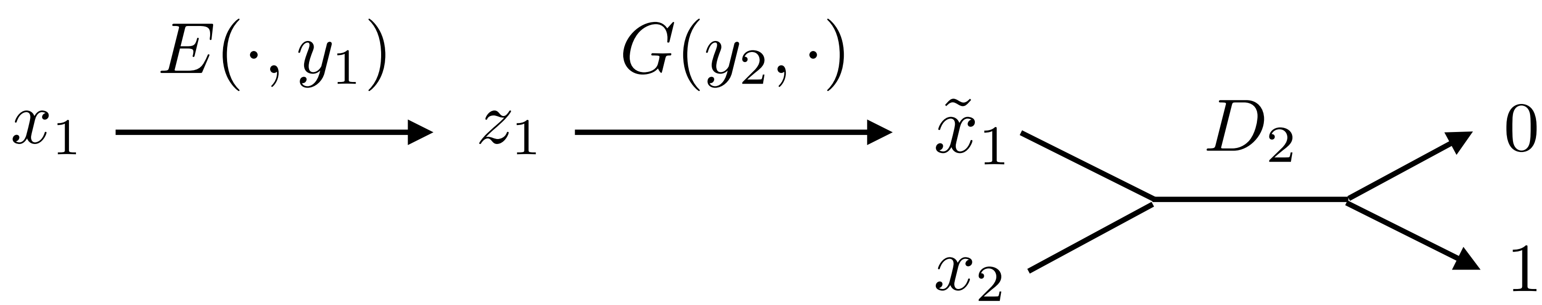
Introduce two discriminators:

D_1 tries to distinguish x_1 and transferred x_2

D_2 tries to distinguish x_2 and transferred x_1

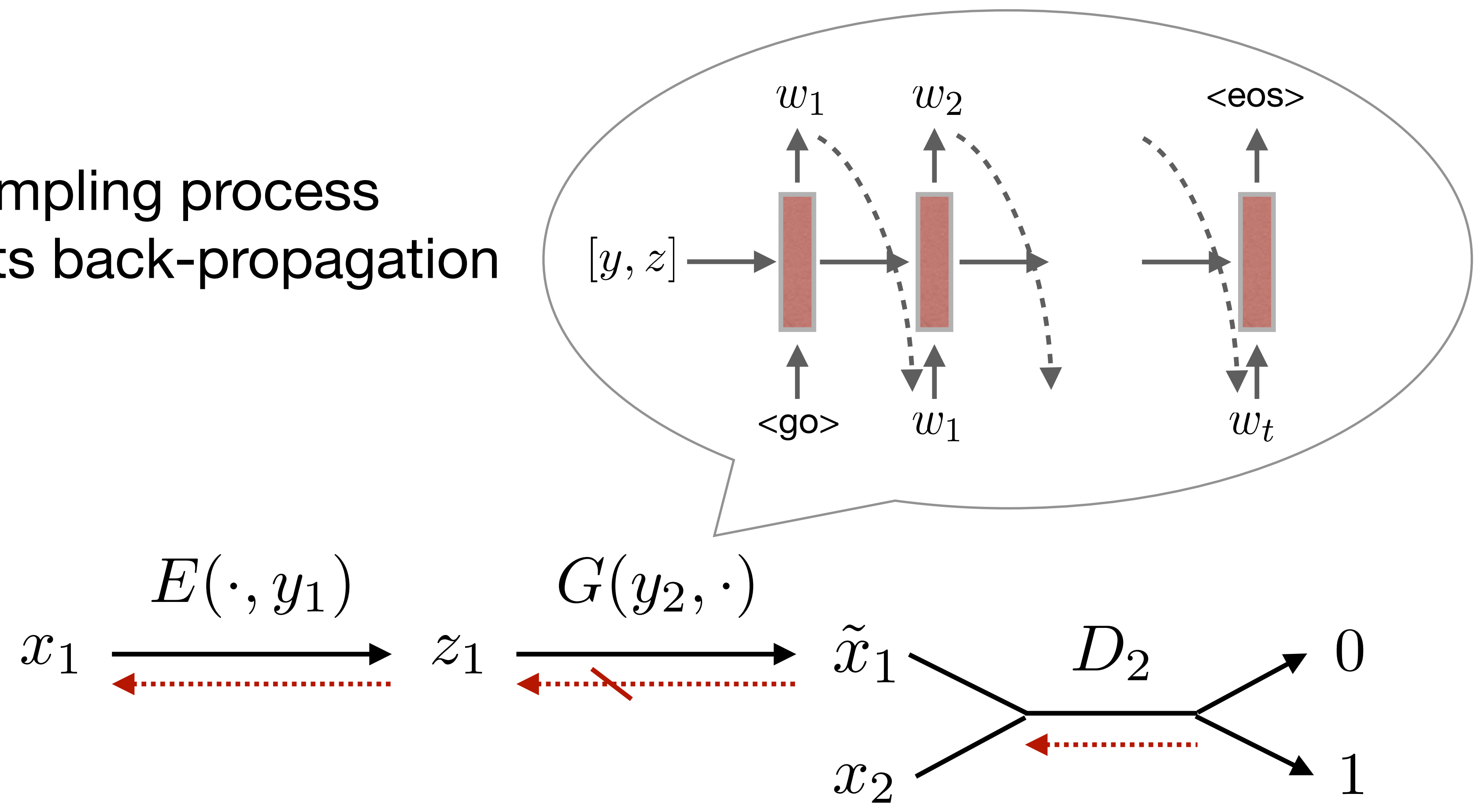
$$\boxed{\min_{E,G} \max_{D_1, D_2} \mathcal{L}_{\text{rec}} - \lambda(\mathcal{L}_{\text{adv}_1} + \mathcal{L}_{\text{adv}_2})}$$

Cross Alignment

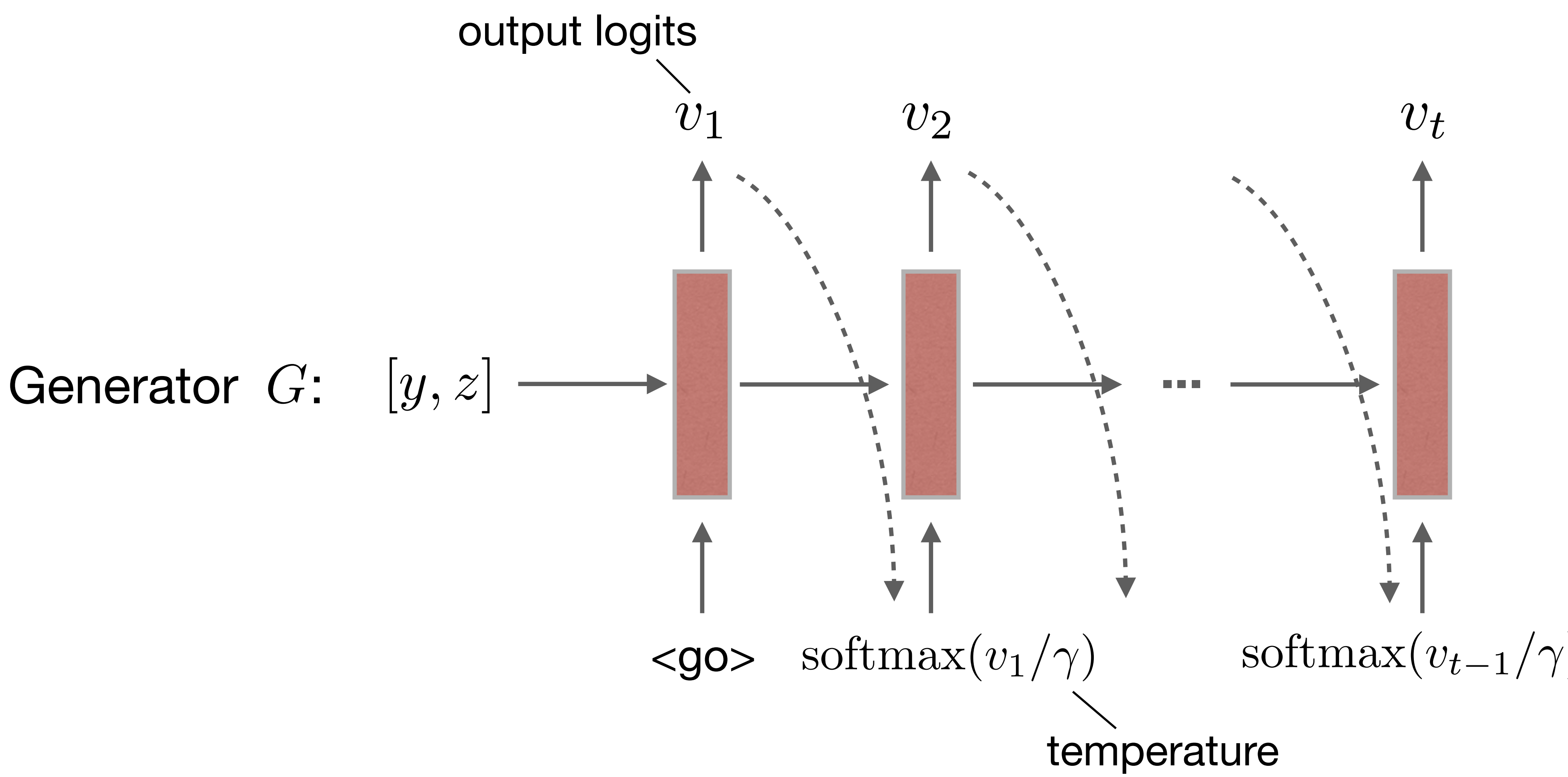


Cross Alignment

discrete sampling process
hinders gradients back-propagation

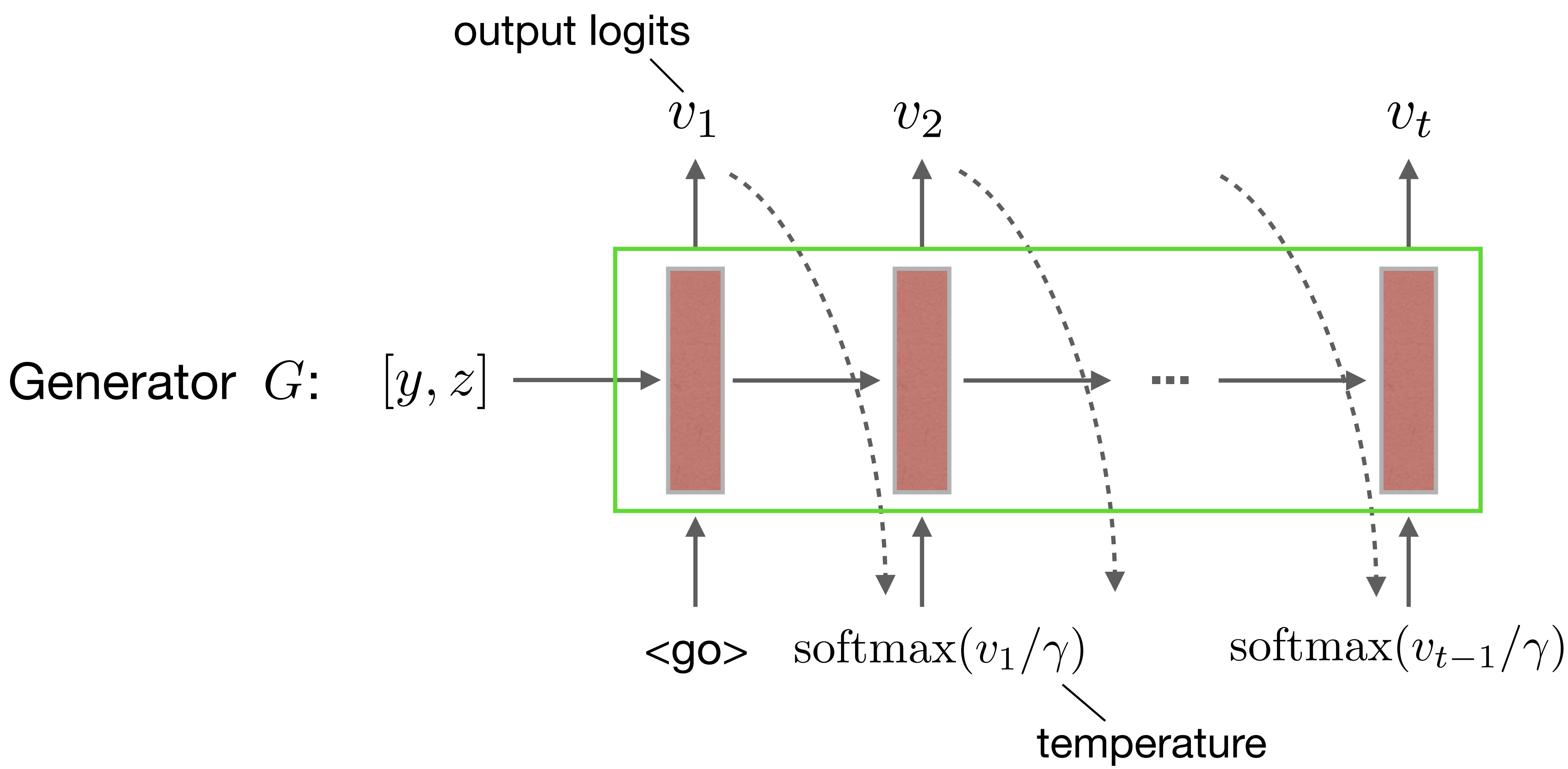


Continuous Relaxation



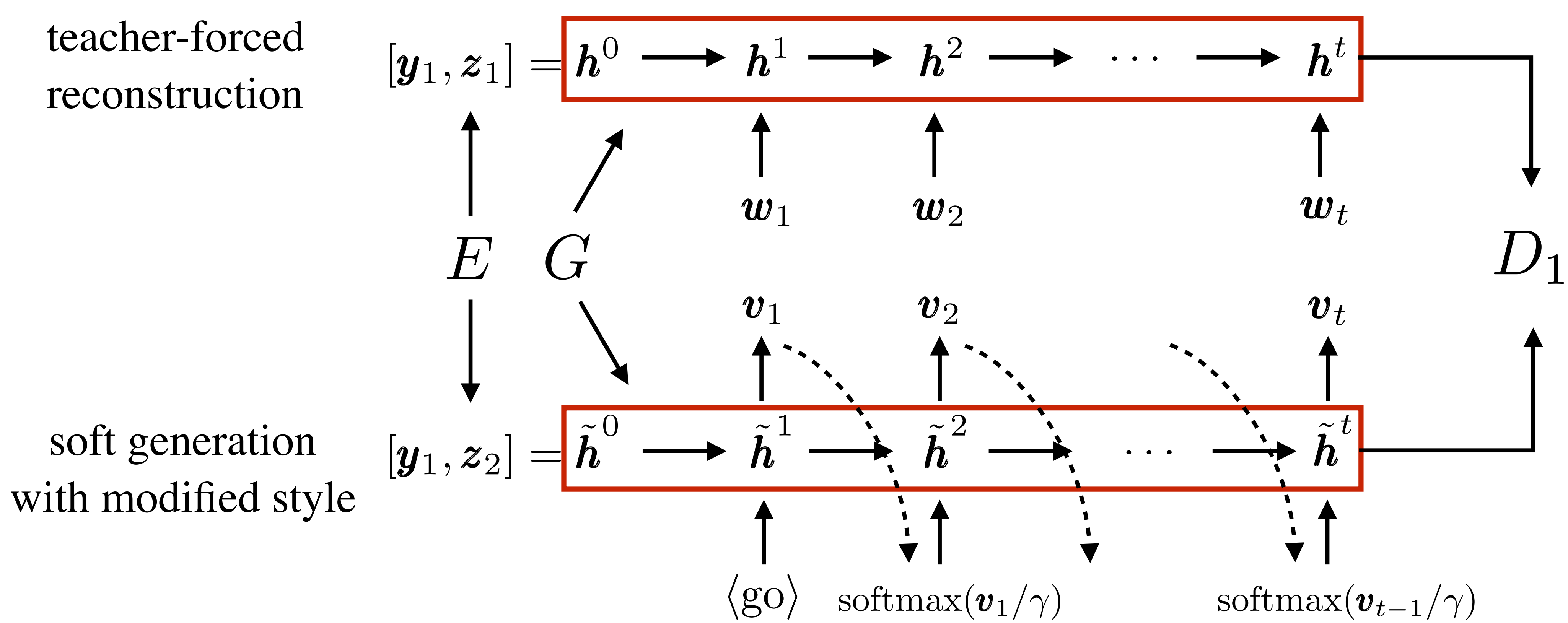
Professor Forcing

Match hidden states instead of output words
- contain all the information, smoothly distributed



[Lamb et al. 2016]

Cross-Aligned Auto-Encoder



Cross-aligning between x_1 and transferred x_2

Enhances aligned auto-encoder, where only the first hidden states z_1 and z_2 are aligned

Cross-Aligned Auto-Encoder

Training procedure:

Take two mini-batches $\{x_1^{(i)}\}_{i=1}^k$ from X_1 and $\{x_2^{(i)}\}_{i=1}^k$ from X_2

- Encode with E $\longrightarrow$ $z_1^{(i)}, z_2^{(i)}$
- Unroll G from $(y_1, z_1^{(i)}), (y_2, z_2^{(i)})$ $\longrightarrow$ $h_1^{(i)}, h_2^{(i)}$
  (reconstruction, teacher-forced by $x^{(i)}$)
- Unroll G from $(y_2, z_1^{(i)}), (y_1, z_2^{(i)})$ $\longrightarrow$ $\tilde{h}_1^{(i)}, \tilde{h}_2^{(i)}$
  (style transfer, self-fed by previous output logits)

Update D_1 (and symmetrically D_2) by gradient descent on loss:

$$\mathcal{L}_{\text{adv}_1} = -\frac{1}{k}\sum_{i=1}^{k} \log D_1(h_1^{(i)}) - \frac{1}{k}\sum_{i=1}^{k} \log(1 - D_1(\tilde{h}_2^{(i)}))$$

Update E, G by gradient descent on loss $\mathcal{L}_{\text{rec}} - \lambda(\mathcal{L}_{\text{adv}_1} + \mathcal{L}_{\text{adv}_2})$

Cross-Aligned Auto-Encoder

Results:

great !
horrible !

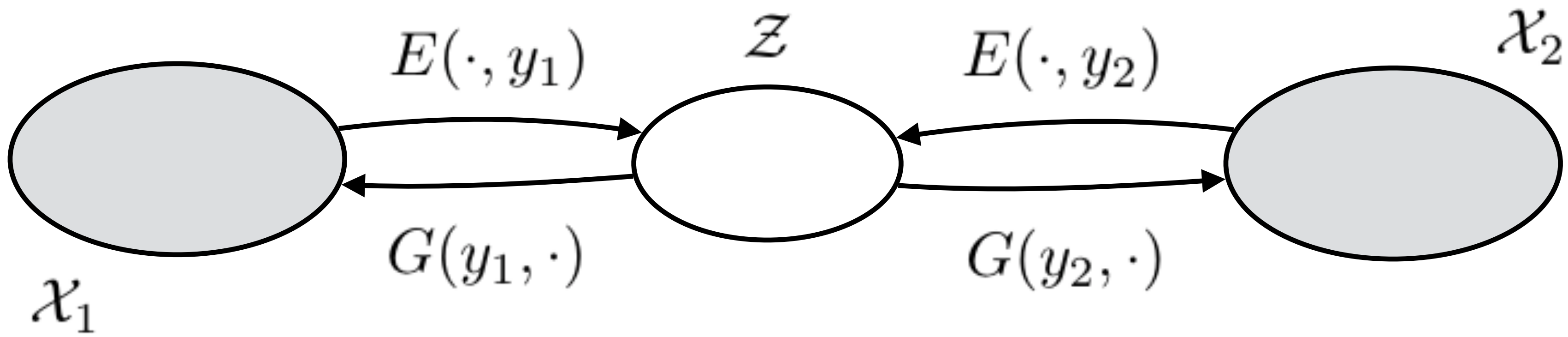
mediocre dim sum if you 're from southern california .
good dim sum if you have korean friends .

i would n't bother .
i would recommend !

i would never go back for the food .
i would definitely go back for the food .

- 78.4% sentiment accuracy as measured by a classifier

Variational Auto-Encoder



Impose a prior $p(z) \sim \mathcal{N}(0, I)$

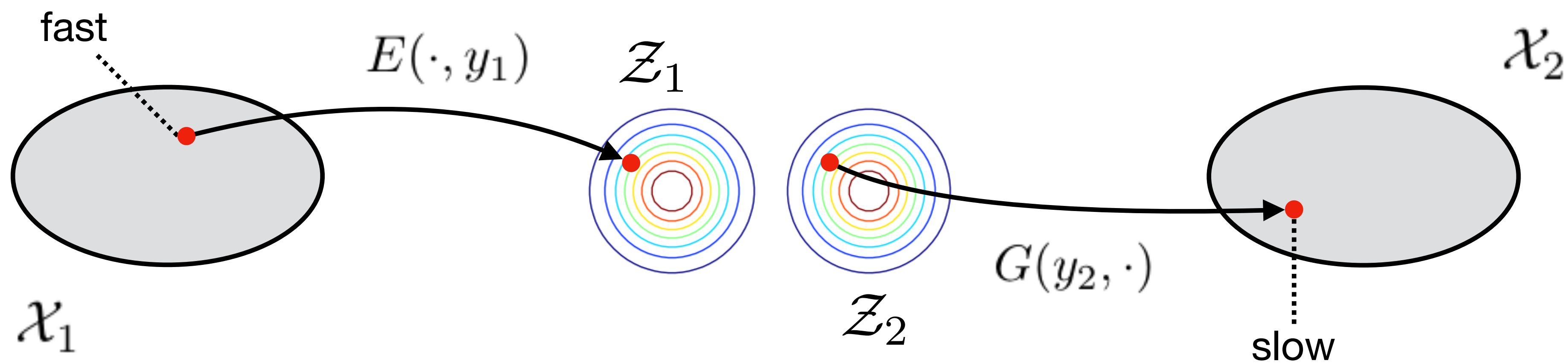
Maximize variational lower bound of data likelihood $-(\mathcal{L}_{\text{rec}} + \mathcal{L}_{\text{KL}})$

$$\begin{aligned}\mathcal{L}_{\text{KL}}(\theta_E) = \; & \mathbb{E}_{x_1 \sim X_1}[D_{\text{KL}}(p_E(z|x_1, y_1) \| p(z))] \; + \\ & \mathbb{E}_{x_2 \sim X_2}[D_{\text{KL}}(p_E(z|x_2, y_2) \| p(z))]\end{aligned}$$

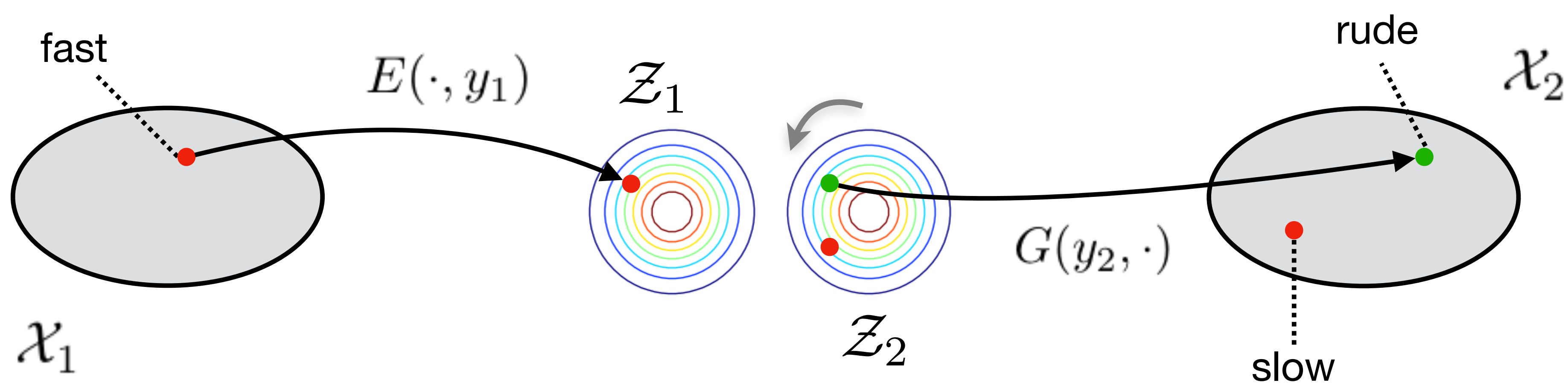
Align both posteriors to the prior

[Kingma and Welling 2013]

Variational Auto-Encoder



Variational Auto-Encoder



Distributional alignment $\xrightarrow{?}$ instance-level matching

Limiting z to a simple and even distribution is detrimental to content preservation

Sentiment Transfer Results

Model Evaluation

Method	accuracy
Hu et al. (2017)	83.5
Variational auto-encoder	23.2
Aligned auto-encoder	48.3
Cross-aligned auto-encoder	78.4

Human Evaluation

Method	sentiment	fluency	overall transfer
Hu et al. (2017)	70.8	3.2	41.0
Cross-align	62.6	2.8	41.5

“Is the transferred sentence semantically equivalent to the source sentence with an opposite sentiment?”

Development of appropriate evaluation measures is crucial

Sentiment Transfer Results

consistently slow .
consistently good .
consistently fast .

my goodness it was so gross .
my husband 's steak was phenomenal .
my goodness was so awesome .

i love the ladies here !
i avoid all the time !
i hate the doctors here !

came here with my wife and her grandmother !
came here with my wife and hated her !
came here with my wife and her son .

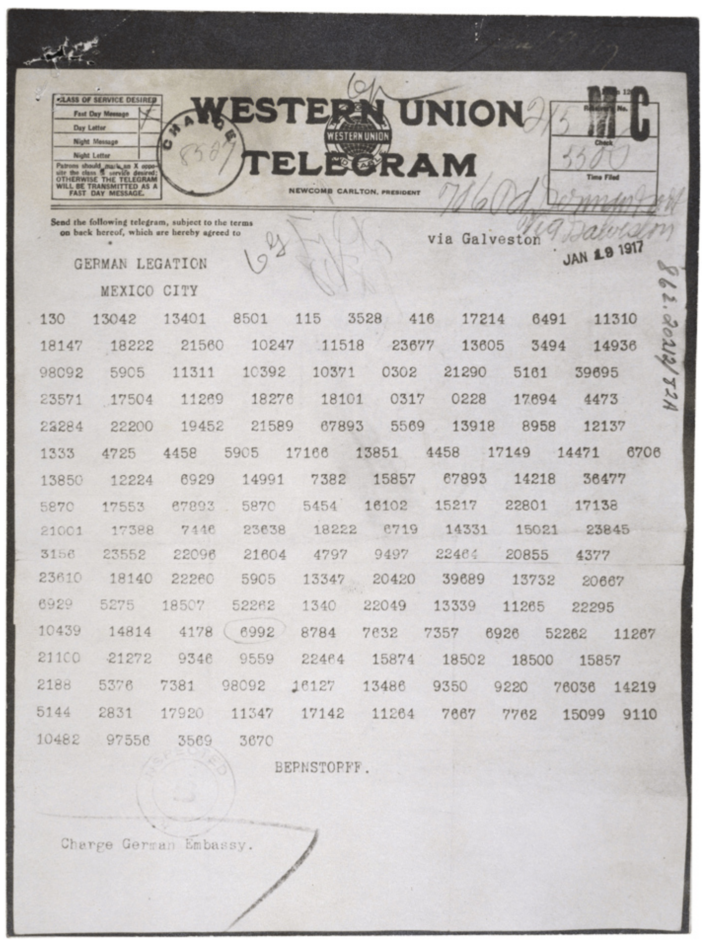
first line—input, second—Hu et al. (2017), third—Cross-align

Decipherment

Non-parallel transfer

Access only to the cipher text, want to transfer it into plain text

Keep the meaning, vary its style



CLASS OF SERVICE DESIRED
Fast Day Message
Day Letter
Night Message
Night Letter
Patrons should mark an X opposite the class of service desired; OTHERWISE THE TELEGRAM WILL BE TRANSMITTED AS A FAST DAY MESSAGE.

WESTERN UNION TELEGRAM

NEWCOMB CARLTON, PRESIDENT

Receiver's No.
Check
Time Filed

Send the following telegram, subject to the terms on back hereof, which are hereby agreed to

via Galveston

JAN 19 1917

GERMAN LEGATION

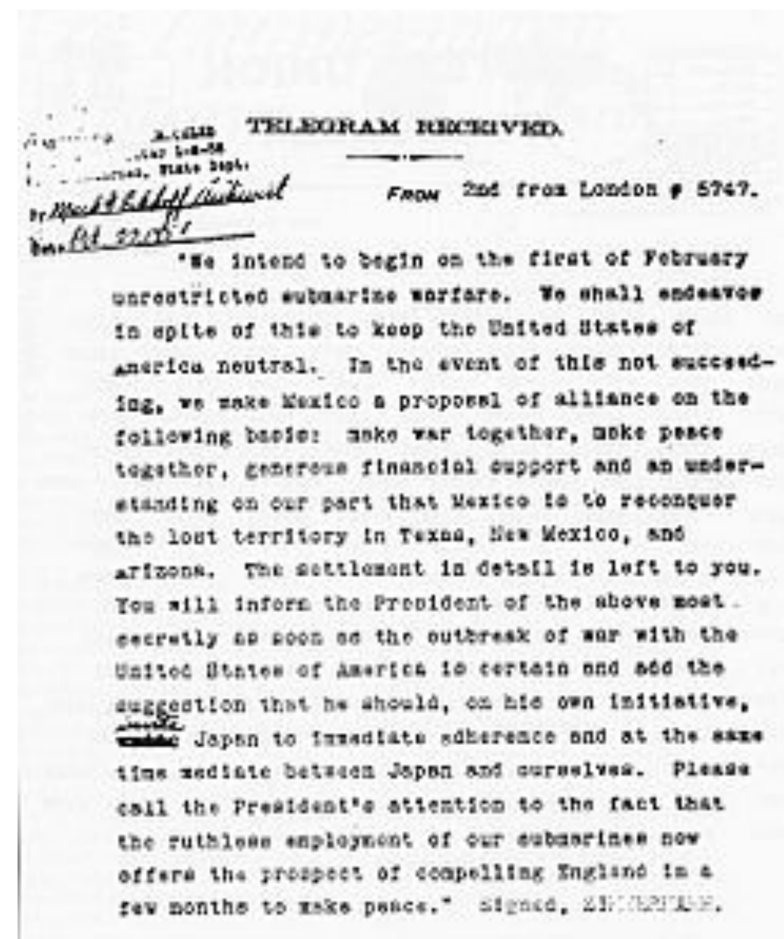
MEXICO CITY

130 13042 13401 8501 115 3528 416 17214 6491 11310
18147 18222 21560 10247 11518 23677 13605 3494 14936
98092 5905 11311 10392 10371 0302 21290 5161 39695
23571 17504 11269 18276 18101 0317 0228 17694 4473
22284 22200 19452 21589 67893 5569 13918 8958 12137
1333 4725 4458 5905 17166 13851 4458 17149 14471 6706
13850 12224 6929 14991 7382 15857 67893 14218 36477
5870 17553 67893 5870 5454 16102 15217 22801 17138
21001 17388 7446 23638 18222 6719 14331 15021 23845
3156 23552 22096 21604 4797 9497 22464 20855 4377
23610 18140 22260 5905 13347 20420 39689 13732 20667
6929 5275 18507 52262 1340 22049 13339 11265 22295
10439 14814 4178 6992 8784 7632 7357 6926 52262 11267
21100 21272 9346 9559 22464 15874 18502 18500 15857
2188 5376 7381 98092 16127 13486 9350 9220 76036 14219
5144 2831 17920 11347 17142 11264 7667 7762 15099 9110
10482 97556 3569 3670

BERNSTORFF.

Charge German Embassy.

cipher text



TELEGRAM RECEIVED.

FROM 2nd from London # 5747.

"We intend to begin on the first of February unrestricted submarine warfare. We shall endeavor in spite of this to keep the United States of America neutral. In the event of this not succeeding, we make Mexico a proposal of alliance on the following basis: make war together, make peace together, generous financial support and an understanding on our part that Mexico is to reconquer the lost territory in Texas, New Mexico, and Arizona. The settlement in detail is left to you. You will inform the President of the above most secretly as soon as the outbreak of war with the United States of America is certain and add the suggestion that he should, on his own initiative, invite Japan to immediate adherence and at the same time mediate between Japan and ourselves. Please call the President's attention to the fact that the ruthless employment of our submarines now offers the prospect of compelling England in a few months to make peace." Signed, [illegible].

plain text

Word Substitution Decipherment

Map every word to a cipher token according to a 1-to-1 substitution key

cipher text	**plain text**
eht azzip saw ton doog	*the pizza was not good*
ew lliw ton eb kcab	*we will not be back*
doog remotsuc ecivres	*good customer service*
os ytsan	*so nasty*
ym ssendoog ti saw os ssorg	*my goodness it was so gross*
ym etirovaf azzip	*my favorite pizza*

Word Substitution Decipherment

Non-parallel training, parallel evaluation

Method	Substitution decipher 20%	40%	60%	80%	100%
No transfer (copy)	56.4	21.4	6.3	4.5	0
Unigram matching	74.3	48.1	17.8	10.7	1.2
Variational auto-encoder	79.8	59.6	44.6	34.4	0.9
Aligned auto-encoder	81.0	68.9	50.7	45.6	7.2
Cross-aligned auto-encoder	**83.8**	**79.1**	**74.7**	**66.1**	**57.4**
Parallel translation	99.0	98.9	98.2	98.5	97.2

Bleu score between plain text and transferred cipher text

Word Ordering

Randomly shuffle a sentence, recover its original word order

bag of words	**grammatical sentence**
! 'm i impressed so	*i 'm so impressed !*
was ca even it how i . gross handle n't	*i ca n't even handle how gross it was .*
really . is which they have good and daily also ice specials cream	*they also have daily specials and ice cream which is really good .*

Word Ordering

Non-parallel training, parallel evaluation

Method	Order recover
No transfer (copy)	5.1
Variational auto-encoder	5.3
Aligned auto-encoder	5.2
Cross-aligned auto-encoder	**26.1**
Parallel translation	64.6

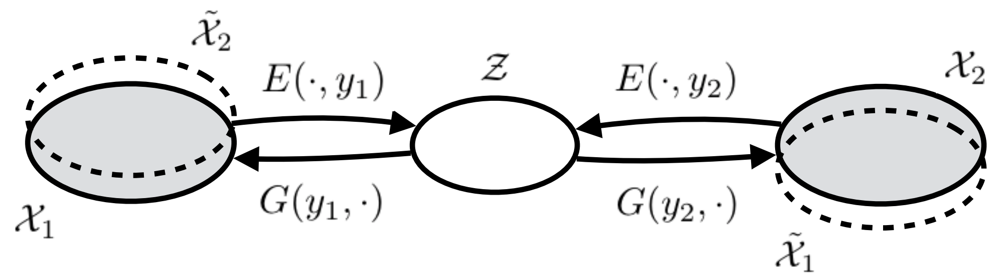
Bleu score between grammatical sentences and transferred shuffled sentences

Conclusion

- Non-parallel style transfer

  keep the content, vary the style

- Cross-aligned auto-encoder



  transferred sentences from one
  style should match example sentence from the other style

- Distributional alignment $\xrightarrow{?}$ instance-level matching

- Applications

  sentiment transfer, decipherment, word ordering

Future Work

- *Real* language style transfer

  critic ⟷ general audience movie reviews

  Shakespeare ⟷ Trump, CNN ⟷ Fox news

- Evaluation

  how to measure the transferred sentence preserves the content?

  how to measure it has the target style?

- Better model

  attention, specific constraints…

Paper: Tianxiao Shen, Tao Lei, Regina Barzilay, and Tommi Jaakkola. Style Transfer from Non-Parallel Text by Cross-Alignment. *NIPS 2017.*

Code & data: https://github.com/shentianxiao/language-style-transfer