Matching Structure for Dual Learning

ICML 2022

Hao Fei, Shengqiong Wu, Yafeng Ren, Meishan Zhang

Sea-NExT Joint Lab, National University of Singapore, Singapore
Guangdong University of Foreign Studies, China
Harbin Institute of Technology (Shenzhen), China

● Motivation

➢ Dual Learning

✓ Many *NLP/CV/Multimodal* tasks appear in dual forms.

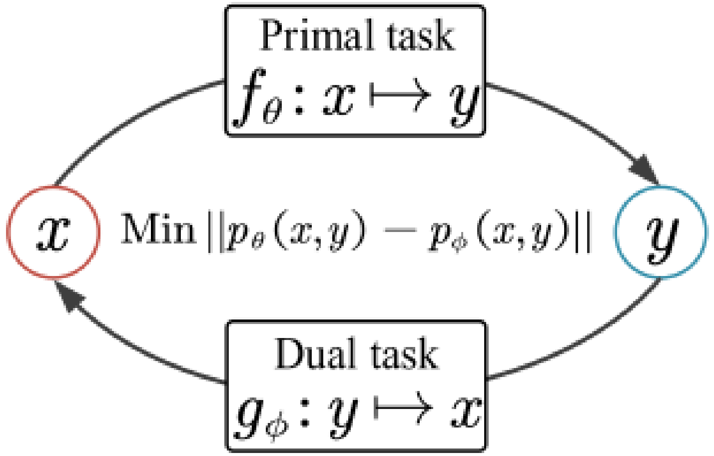
- The primal and dual tasks have the same exact input and output but in reverse.

Duality Scheme	Direction	Representative Application(s)
Text↔Text	$\longrightarrow$ or $\longleftarrow$	Neural Machine Translation, Paraphrase Generation
Text↔Image	$\longrightarrow$ $\longleftarrow$	Text-to-Image Synthesis Image Captioning
Text↔Label	$\longrightarrow$ $\longleftarrow$	Text Classification Conditioned Text Generation
Image↔Label	$\longrightarrow$ $\longleftarrow$	Image Classification Conditioned Image Generation
Image↔Image	$\longrightarrow$ or $\longleftarrow$	Image Translation

✓ Dual learning scheme

- Modeling the duality between the task pair, by minimizing the gap between joint distributions of the two tasks respectively.

$$
\begin{aligned}
& p_\theta(x,y) = p(x)p(y|x;\theta) \\
\simeq \quad & p_\phi(x,y) = p(y)p(x|y;\phi)\,, \forall x \& y\,,
\end{aligned}
$$



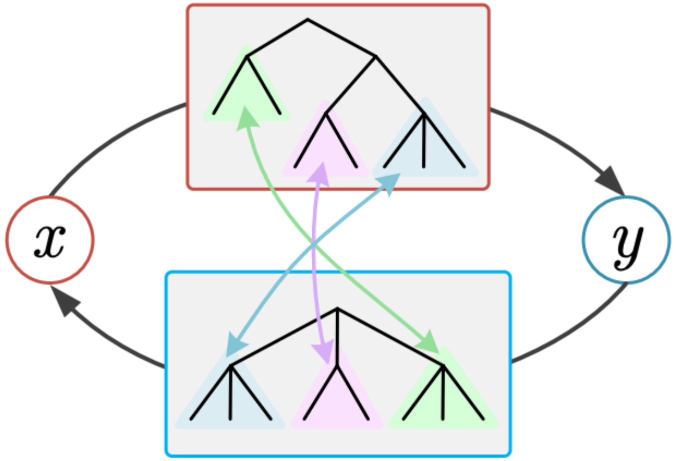
● Motivation

➢ Existing Problem

✓ Current dual learning fails to explicitly model the **structural correspondence** between two coupled tasks.

✓ Structure features are important to many learning tasks:

- neural machine translation
- paraphrase generation
- conditioned text generation
- …



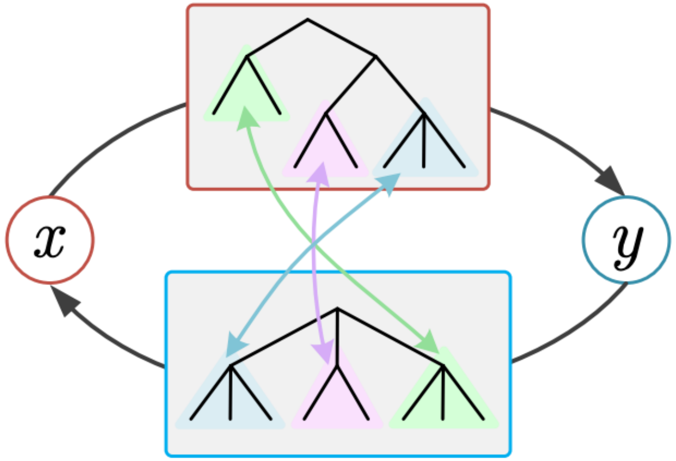
● Method

➢ Our proposal

◆ Matching Structure for Dual Learning

✓ *Core idea:*

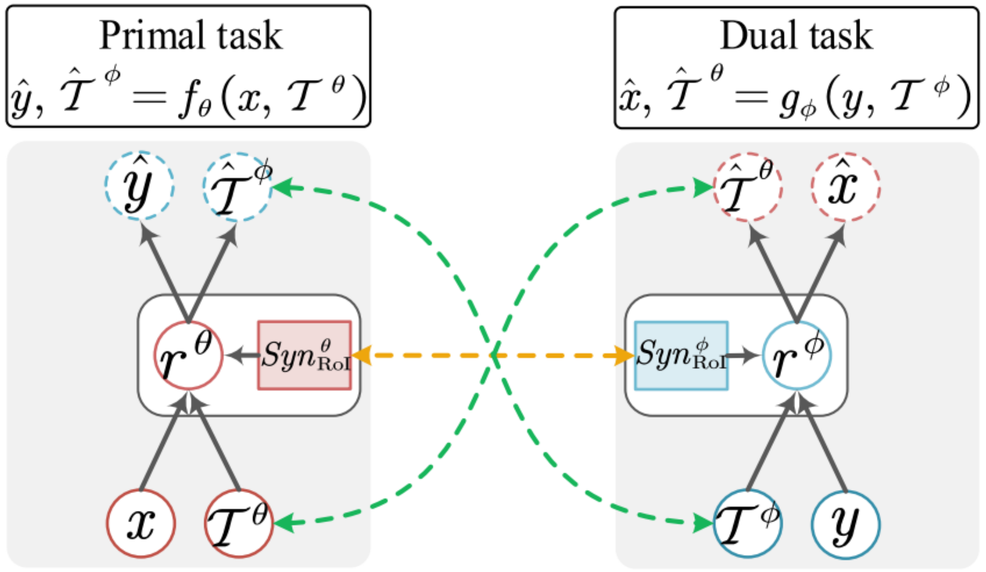
Based on the vanilla, dual learning framework, we perform structural alignment unsupvervisedly between the primal and dual tasks, bridging them with structure connections.



● Method

➢ **Dually-Syntactic Structure Matching** for **Text ↔ Text Dual Learning**

- Symmetrically syntactic structure matching for dual learning



$$\mathcal{L}(\theta,\phi) = \boxed{\mathcal{L}_C} + \boxed{\lambda_1\mathcal{L}_D} + \lambda_2\mathcal{L}_M + \lambda_3\mathcal{L}_R$$

- Task learning of two coupled tasks

$$\mathcal{L}_\theta = \mathbb{E}_{x,y}\ \log p(y|x;\theta)\ ,$$
$$\mathcal{L}_\phi = \mathbb{E}_{x,y}\ \log p(x|y;\phi)\ .$$
$$\mathcal{L}_C = \mathcal{L}_\theta + \mathcal{L}_\phi .$$

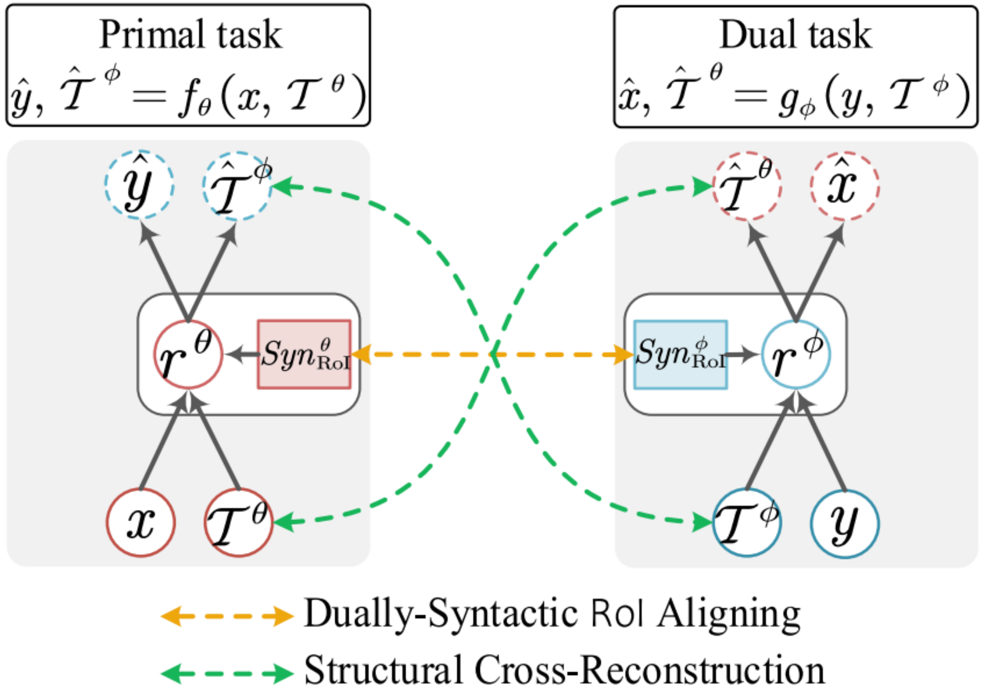
- Dual learning backbone

$$\mathcal{L}_D = || \log \hat{p}(x) + \log p(y|x;\theta) - \log \hat{p}(y) - \log p(x|y;\phi)||\ ,$$

Method

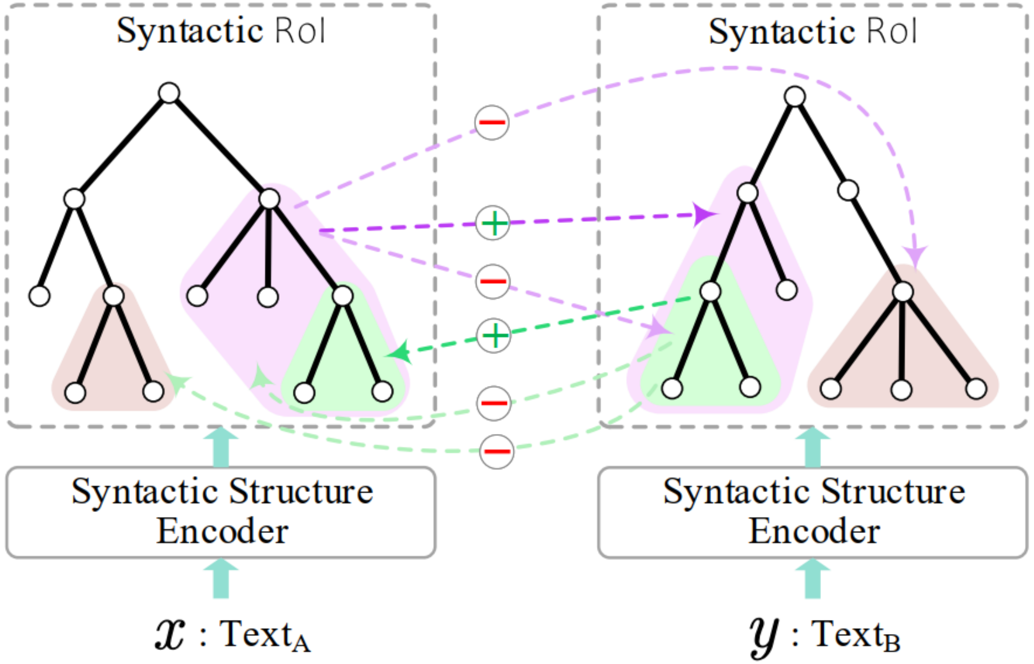
➢ Dually-Syntactic Structure Matching for Text ↔ Text Dual Learning

- Symmetrically syntactic structure matching for dual learning



$$\mathcal{L}(\theta, \phi) = \mathcal{L}_C + \lambda_1 \mathcal{L}_D + \boxed{\lambda_2 \mathcal{L}_M} + \lambda_3 \mathcal{L}_R$$

- Dually-syntactic RoI alignment



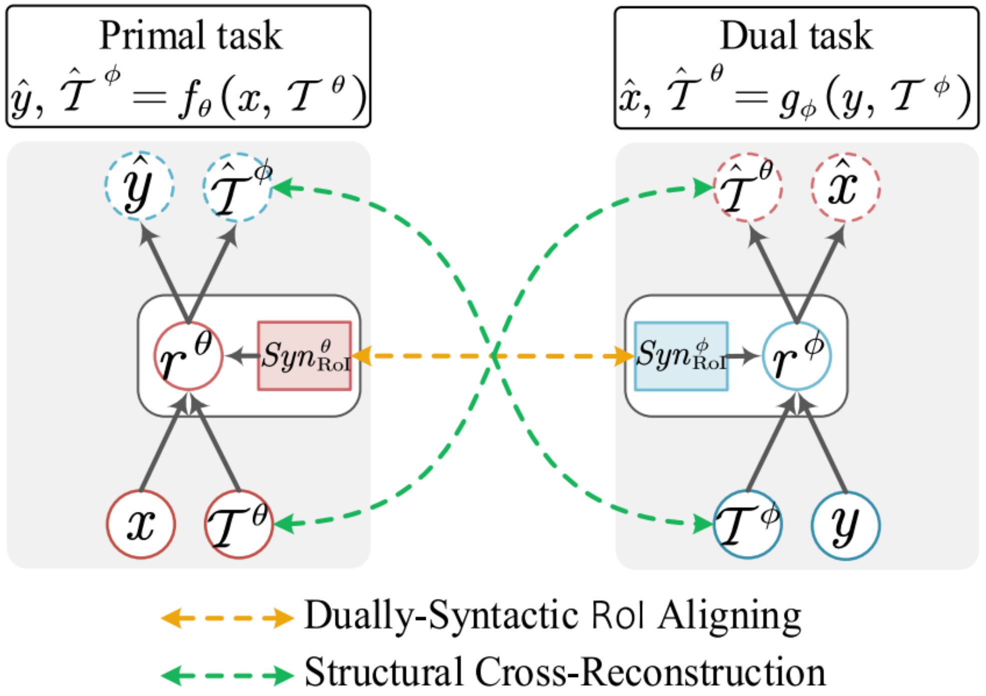
$$\mathcal{L}_M = - \sum_{i \in \mathcal{T}^{\theta},\, j^* \in \mathcal{T}^{\phi}} \log \frac{\exp(s_{i,j^*}/\tau)}{\mathcal{Z}},$$

$$\mathcal{Z} = \sum_{i \in \mathcal{T}^{\theta},\, k \in \mathcal{T}^{\phi},\, k \neq j^*} \exp(s_{i,k}/\tau),$$

● Method

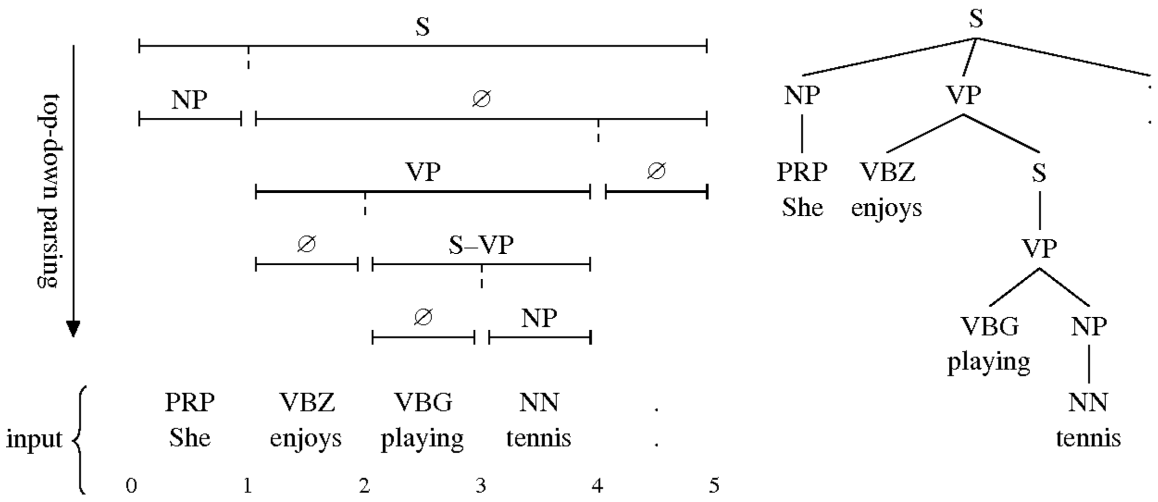
➢ **Dually-Syntactic Structure Matching** for **Text ↔ Text Dual Learning**

- Symmetrically syntactic structure matching for dual learning



$$\mathcal{L}(\theta,\phi) = \mathcal{L}_C + \lambda_1\mathcal{L}_D + \lambda_2\mathcal{L}_M + \boxed{\lambda_3\mathcal{L}_R}$$

- Structural Cross-Reconstruction



$$\mathcal{L}_R = \mathcal{L}_R^{\theta} + \mathcal{L}_R^{\phi} .$$

● Method

➢ Exp-I: Text↔Text Applications

1) Comparing M2 to M1 and M4 to M3:

✓ *the integration of syntactic structure results in better performances, either for the singleton or dual learning*

2) Comparing M3 to M1:

✓ *the dual learning technique improves the task performances consistently*

3) Comparing M4 to OnlySyn:

✓ *high efficacy of the structural matching proposal*

4) Comparing M4-SAln vs. M4-SyRec:

✓ *the RoI alignment mechanism plays the predominant influences than the syntactic structure reconstruction mechanism*

5) Comparing M4(CL) vs. M4(RANK):

✓ *the contrastive learning can bring better effectiveness than the ranking loss method*

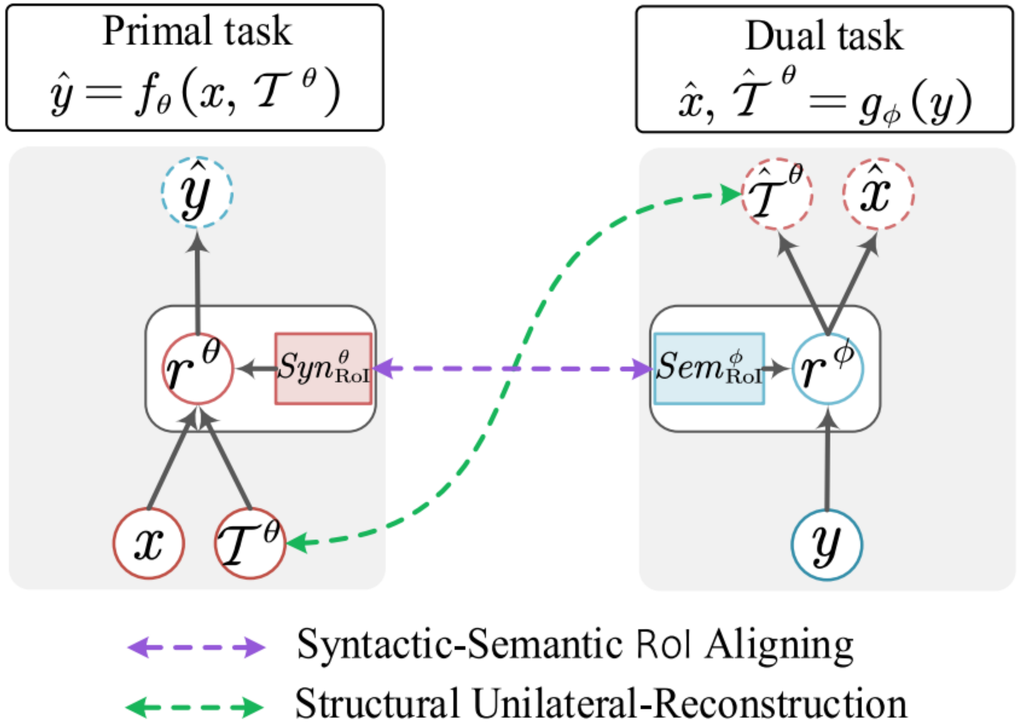
		ParaNMT						QUORA					
		B	R-1	R-2	R-L	B	R-1	B	R-1	R-2	R-L	B	R-1
• *Baseline*	B1	20.4	50.3	25.2	51.6	21.8	46.4	19.5	40.6	22.5	44.6	17.8	44.1
	B2	20.8	49.6	28.4	48.6	19.0	45.0	22.3	56.4	26.2	52.3	21.0	52.8
	B3	23.6	54.8	32.0	58.3	25.4	48.7	30.4	62.6	42.7	65.4	28.1	60.5
	B4	27.5	60.6	36.9	54.5	27.2	53.2	35.8	68.1	45.7	70.2	35.6	65.7
• *Transformer-based*	M1	24.6	50.3	30.7	45.8	25.4	51.7	29.7	58.5	37.5	59.6	28.0	60.5
	M2	27.2	56.4	34.4	50.6	26.1	53.6	33.4	63.4	41.8	63.4	34.8	65.8
	M3	26.2	57.1	33.0	53.5	27.8	55.9	32.0	65.7	40.0	66.4	34.0	64.3
	M4(RANK)	30.1	61.8	38.9	59.8	30.2	62.5	37.3	70.4	47.2	72.4	37.4	71.2
	M4(CL)	30.5	62.4	39.4	60.4	30.6	62.7	37.5	70.5	47.6	72.5	37.5	71.5
	OnlySyn	27.7	58.9	34.9	54.7	28.0	56.2	33.7	66.4	42.0	67.1	35.0	65.8
	-SAln	28.0	59.6	35.8	56.0	28.6	57.3	34.6	67.6	43.2	68.9	35.8	67.4
	-SyRec	29.7	60.2	37.8	58.3	29.7	61.0	36.1	68.9	45.0	71.4	36.5	69.3
	M3+BART	33.8	65.7	41.8	62.8	32.7	64.0	41.5	73.3	49.4	74.2	42.0	71.5
	M4+BART	**36.7**	**66.2**	**43.6**	**64.0**	**34.8**	**64.6**	**43.0**	**74.8**	**52.8**	**76.8**	**43.5**	**72.8**

Table 2. Results on paraphrase generation (SRC→TGT, SRC←TGT). B: BLEU, R-X: ROUGE-X.

● Method

➢ **Syntactic-Semantic Structure Matching** for **text ↔ non-text Dual Learning**

- Unsymmetrically syntactic structure matching for dual learning



$$\mathcal{L}(\theta, \phi) = \mathcal{L}_C + \lambda_1 \mathcal{L}_D + \lambda_2 \mathcal{L}_M + \lambda_3 \mathcal{L}_R$$

- Task learning of two coupled tasks

$$\mathcal{L}_\theta = \mathbb{E}_{x,y} \log p(y|x;\theta) \,,$$
$$\mathcal{L}_\phi = \mathbb{E}_{x,y} \log p(x|y;\phi) \,.$$
$$\mathcal{L}_C = \mathcal{L}_\theta + \mathcal{L}_\phi.$$

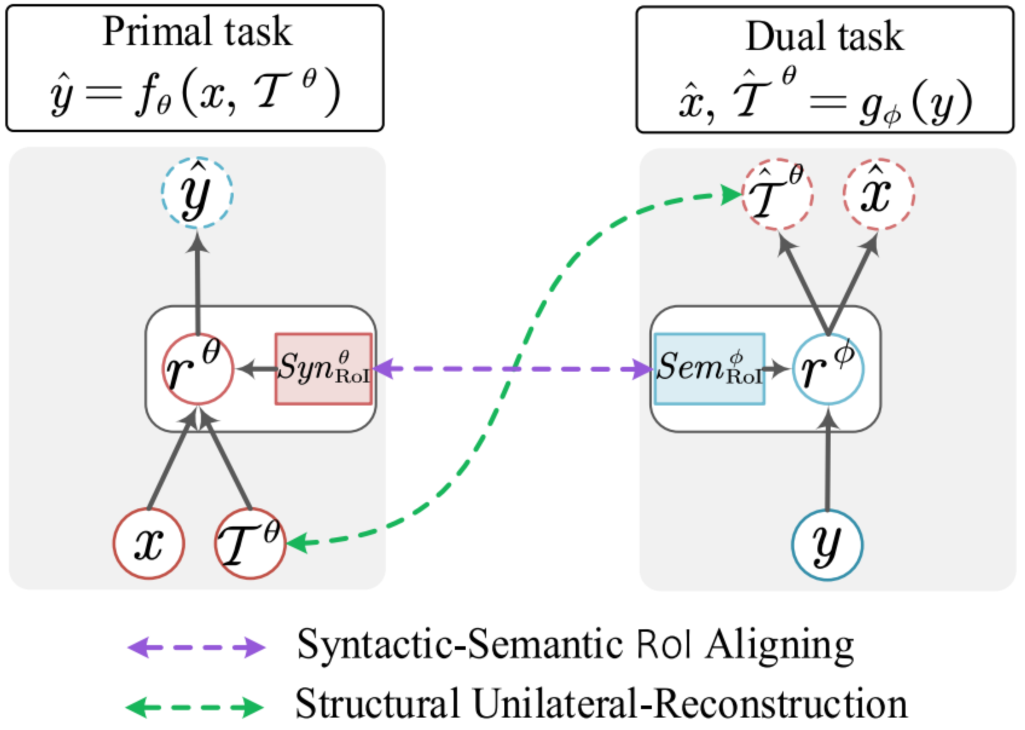
- Dual learning backbone

$$\mathcal{L}_D = || \log \hat{p}(x) + \log p(y|x;\theta) - \log \hat{p}(y) - \log p(x|y;\phi) || \,,$$

● Method

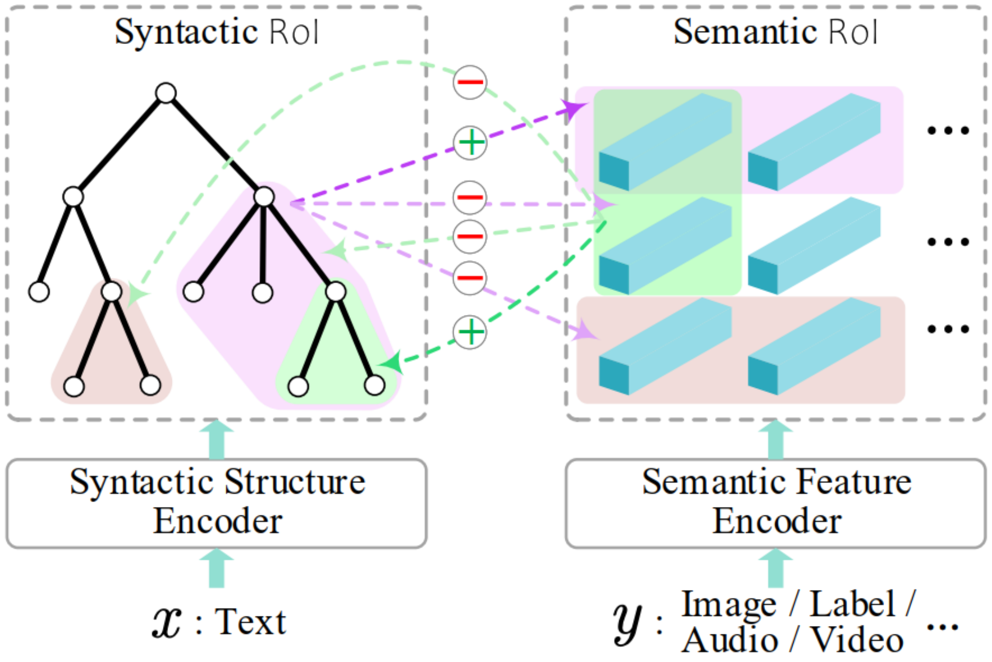
➢ **Syntactic-Semantic Structure Matching** for **text ↔ non-text Dual Learning**

- Unsymmetrically syntactic structure matching for dual learning



$$\mathcal{L}(\theta, \phi) = \mathcal{L}_C + \lambda_1 \mathcal{L}_D + \boxed{\lambda_2 \mathcal{L}_M} + \lambda_3 \mathcal{L}_R$$

- Syntactic-semantic RoI alignment



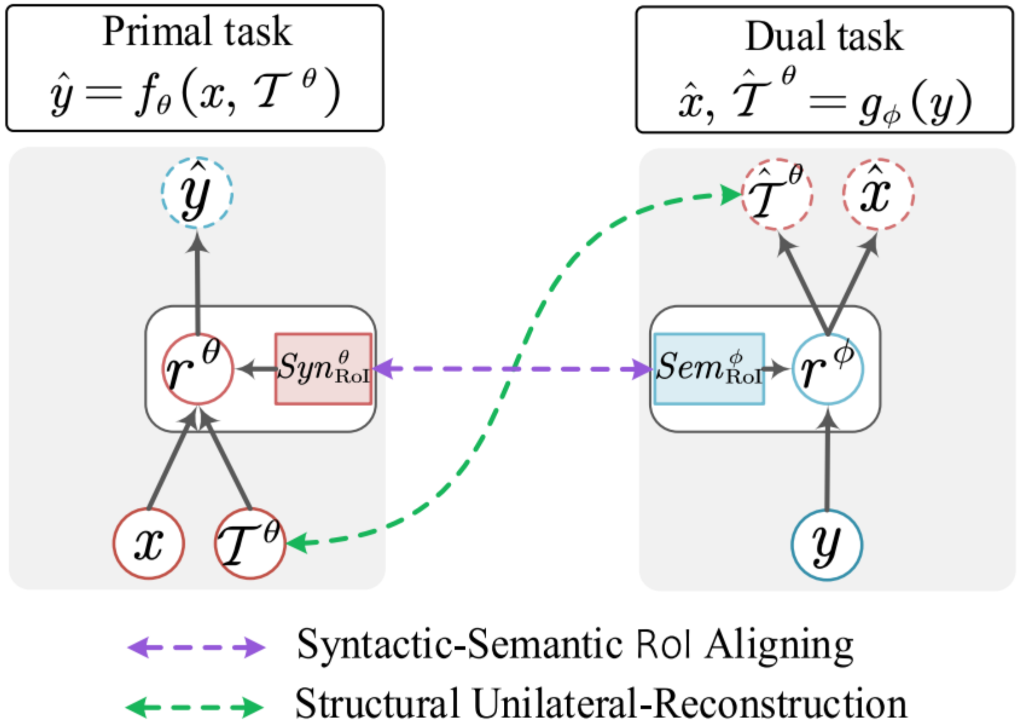
$$\mathcal{L}_M = - \sum_{i \in \mathcal{T}^\theta,\, j^* \in \mathcal{T}^\phi} \log \frac{\exp(s_{i,j^*}/\tau)}{\mathcal{Z}},$$

$$\mathcal{Z} = \sum_{i \in \mathcal{T}^\theta,\, k \in \mathcal{T}^\phi,\, k \neq j^*} \exp(s_{i,k}/\tau),$$

● Method

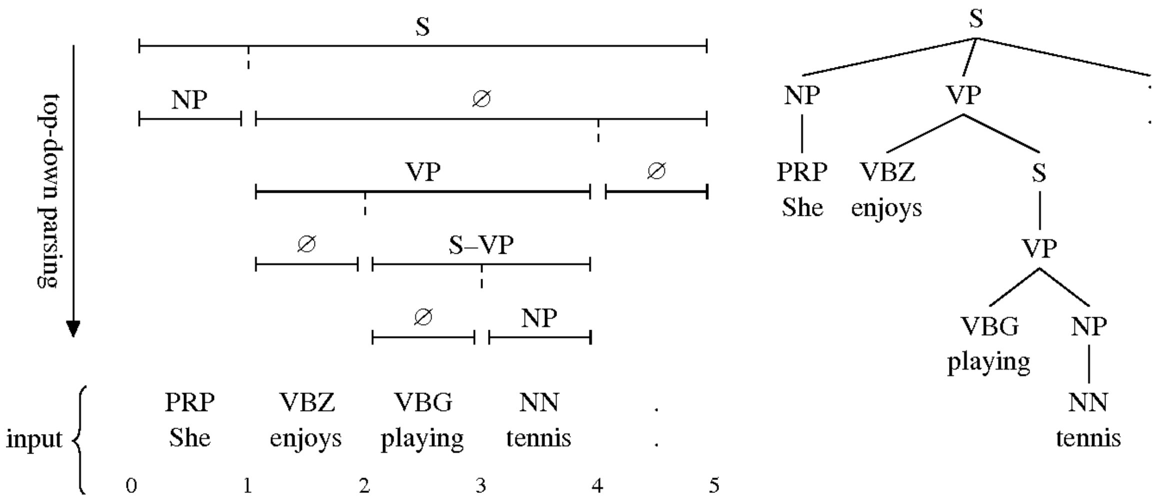
➢ **Syntactic-Semantic Structure Matching** for **text ↔ non-text Dual Learning**

- Unsymmetrically syntactic structure matching for dual learning



$$\mathcal{L}(\theta, \phi) = \mathcal{L}_C + \lambda_1 \mathcal{L}_D + \lambda_2 \mathcal{L}_M + \boxed{\lambda_3 \mathcal{L}_R}$$

- Structural Cross-Reconstruction



$$\mathcal{L}_R = \mathcal{L}_R^\theta + \mathcal{L}_R^\phi .$$

● Method

➢ Exp-II: Text↔Non-Text Applications

	MsCoCo				Flickr30k			
	IS↑	FID↓	B-4	MTR	IS↑	FID↓	B-4	MTR
M1	25.6	28.3	32.5	22.8	6.8	36.8	17.6	15.5
M2	27.8	25.5	/	/	7.5	35.0	/	/
M3	28.4	24.8	36.1	25.1	7.3	34.2	20.1	17.2
M4	**30.7**	**20.6**	**40.0**	**29.6**	**8.0**	**30.9**	**22.6**	**19.5**
-SALN	29.0	21.5	37.3	28.3	7.4	33.0	21.3	17.9
-SYREC	29.8	21.3	39.2	29.0	7.7	31.8	21.9	18.6

Table 3. Results on text↔image experiment (TXT→IMG: text-to-image synthesis, TXT←IMG: image captioning). B-4: BLEU-4, MTR: METEOR.

	Yelp2014				IMDB			
	ACC	B-4	MTR	ACC	ACC	B-4	MTR	ACC
M1	60.6	17.8	33.0	53.8	50.6	17.6	36.9	43.6
M2	61.8	/	/	/	51.9	/	/	/
M3	62.0	19.4	36.4	56.6	53.8	18.3	41.4	47.3
M4	**63.8**	**21.8**	**40.8**	**62.4**	**55.6**	**20.2**	**47.1**	**50.9**
-SALN	63.2	19.9	37.0	57.2	54.2	18.9	44.6	48.4
-SYREC	62.9	20.4	38.5	61.8	55.0	19.5	46.0	49.3

Table 4. Results on Text↔Label experiment (TXT→LB: text classification, TXT←LB: conditioned text generation).

✓ Similar trends with that in the Exp-I: the success of our proposed method can be inherited to the dual learning scenarios more than purely texts.

● Analysis

➢ Four pivotal questions

Questions

★ First, how does structure matching strategy improve the dual learning?

★ Second, for the text generation what are improved when aligning the structures?

★ Third, can the success of the structure alignment be extented to fully non-text scenarios?

★ Fourth, what are the key factors to the structure matching for dual learning?

Analysis

➢ Evaluating correctness of unsupervised structure matching

	WMT14 (EN-DE)		WMT14 (EN-FR)	
	EN→DE	EN←DE	EN→FR	EN←FR
+ Auto RoI	29.03	31.96	41.82	36.76
+ Gold RoI	**29.51**	**32.23**	**42.03**	**36.98**
Δ	-0.48	-0.27	-0.31	-0.22
	ParaNMT		QUORA	
	SRC→TGT	SRC←TGT	SRC→TGT	SRC←TGT
+ Auto RoI	31.53	30.60	38.66	37.58
+ Gold RoI	**31.86**	**30.85**	**39.02**	**38.11**
Δ	-0.33	-0.25	-0.36	-0.53

Table 5. Results (BLEU) of dual learning with automatically learned and gold RoI matching respectively.

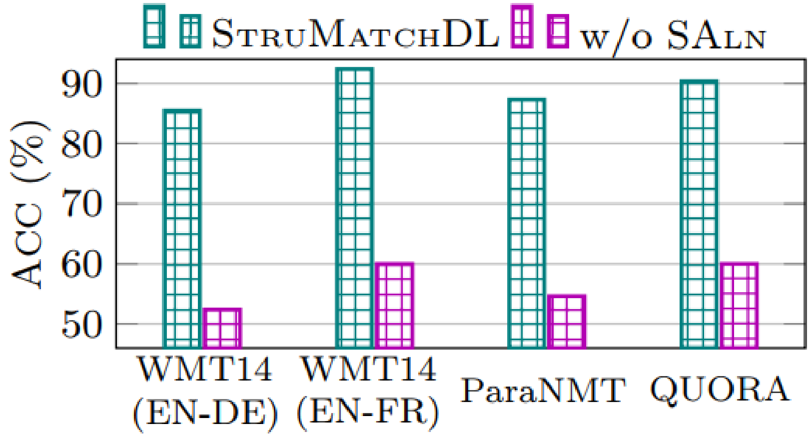


Figure 6. Measuring text↔text RoI alignment.

	ACC
MAF	61.4
StruMatchDL	**54.3 ± 0.3**
-SyRec	46.7 ± 0.5
-SAln	28.6 ± 0.8

Table 6. Visual grounding results on Flickr30k test set for verifying text↔image matching. MAF is a supervised visual grounding system (Wang et al., 2020).

✓ *Structure matching helps correctly retrieve and emphasize the key RoIs that are crucial to the task improvements.*

● Analysis

➢ **Evaluating correctness of unsupervised structure matching**

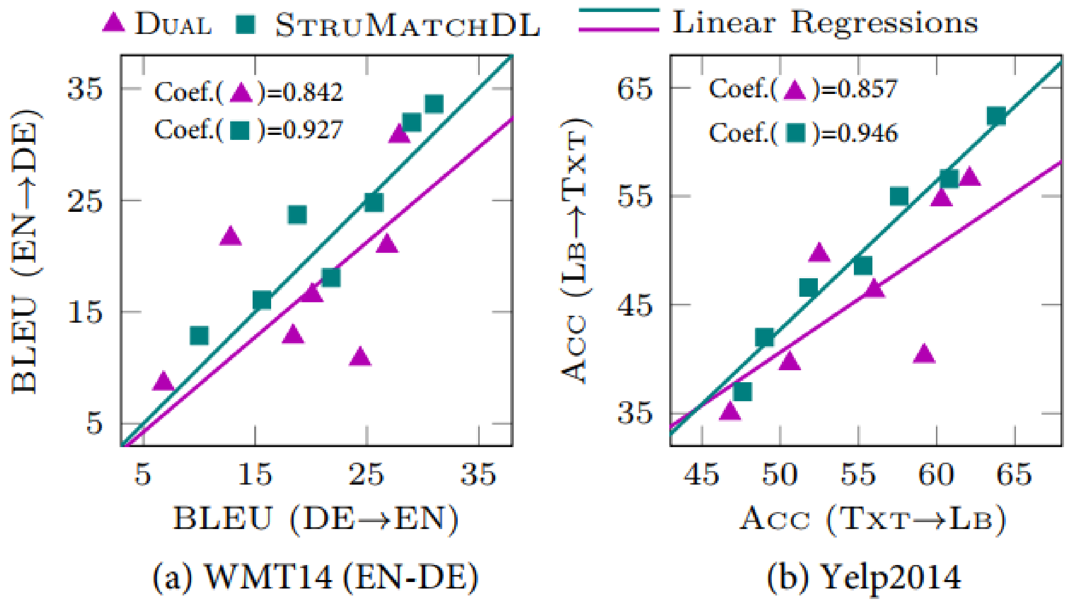


Figure 7. Performance correlation between two coupled tasks. 'Coef.' indicates Pearson correlation coefficient.

✓ *Our method strengthens the duality between two dual tasks by correctly aligning the RoIs.*

Analysis

➢ Evaluating Generated Text

	ParaNMT			MsCoCo		
	Gram.	Corr.	Cont.	Gram.	Corr.	Cont.
HUMAN	4.86	4.92	3.78	4.82	4.15	4.37
BASELINE	1.58	2.20	1.04	0.78	1.23	0.98
DUAL	2.24	2.55	1.46	1.80	2.38	1.25
STRUMATCHDL	**3.78***	**3.67***	2.51	**3.46***	**3.27***	2.74
-SYREC	2.89	3.21	**2.90***	2.75	2.89	**2.96***

Table 7. Human evaluation results. Grammaticality (Gram.), correctness (Corr.), and content richness (Cont.) are rated on Likert 5-scale. ∗ indicates significantly better over the variant (p<0.03).

DUAL STRUMATCHDL

		2	3	4	5	6	7	8
NP	DUAL	15.3	10.5	12.6	2.71	1.56	1.20	0.31
	STRUMATCHDL	10.8	8.10	6.20	10.4	8.10	4.52	3.76
VP	DUAL	15.3	12.5	4.64	1.10	0.95	0.40	0.09
	STRUMATCHDL	6.18	9.23	12.2	7.40	9.17	6.35	5.60
PP	DUAL	1.27	0.57	0.62	0.31	0.00	0.00	0.00
	STRUMATCHDL	7.40	4.10	8.20	2.40	1.50	1.02	0.76
SBAR	DUAL	1.32	0.55	0.06	0.14	0.00	0.03	0.00
	STRUMATCHDL	5.40	6.70	3.26	1.45	0.12	0.82	0.65
ADJP	DUAL	0.60	1.14	0.23	0.10	0.00	0.00	0.00
	STRUMATCHDL	3.40	2.10	1.20	0.40	2.50	0.72	0.12
ADVP	DUAL	0.62	0.24	0.24	0.06	0.00	0.00	0.00
	STRUMATCHDL	1.47	2.10	1.56	0.46	0.24	0.15	0.00

Phrase length (word)

Figure 8. Distribution (frequency, %) over different constituency length of phrases in the generated sentences.

✓ *Our method strengthens the duality between two dual tasks by correctly aligning the RoIs.*

Analysis

➢ Exploring Extendibility

	CIFAR-10			CIFAR-100		
	IMG→LB	IMG←LB		IMG→LB	IMG←LB	
	ACC	IS↑	FID↓	ACC	IS↑	FID↓
M1	93.05	8.62	13.53	72.60	9.34	19.63
M3	93.68	9.83	9.80	73.85	13.64	15.72
M4	**94.74**	**10.64**	**7.38**	**74.63**	**14.65**	**13.42**
Δ	**+1.06**	**+0.81**	**-2.42**	**+0.78**	**+1.01**	**-2.30**

Table 10. Image↔Label experiment (IMG→LB: image classification, IMG←LB: conditioned image generation) on CIFAR-10 and CIFAR-100 datasets.

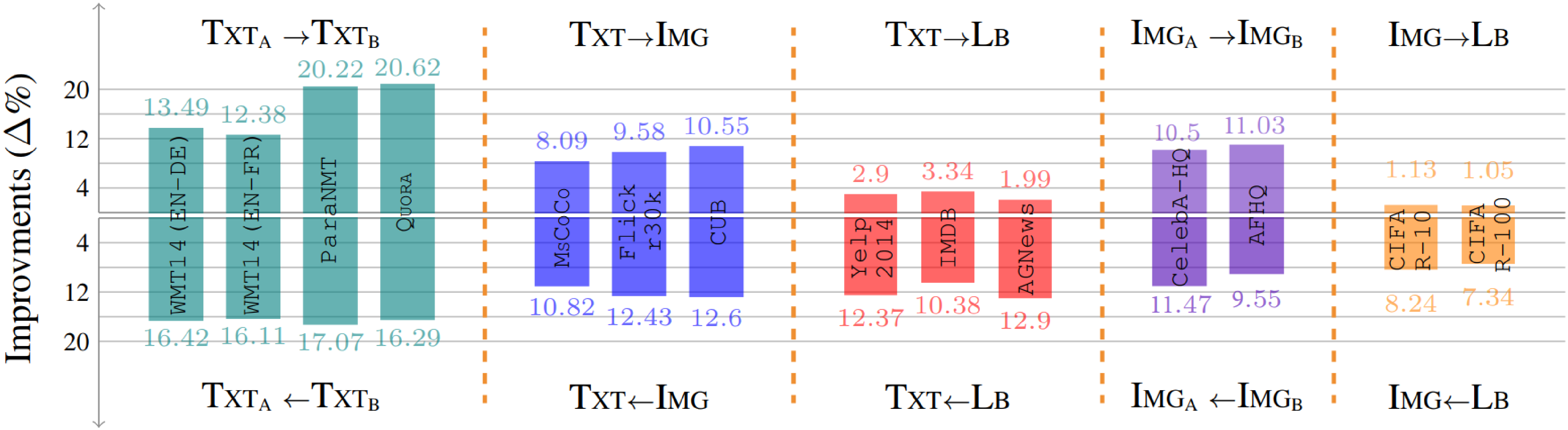
	CelebA-HQ		AFHQ	
	$\text{IMG}_\text{A} \rightarrow \text{IMG}_\text{B}$	$\text{IMG}_\text{A} \leftarrow \text{IMG}_\text{B}$	$\text{IMG}_\text{A} \rightarrow \text{IMG}_\text{B}$	$\text{IMG}_\text{A} \leftarrow \text{IMG}_\text{B}$
M1	26.7	32.7	32.4	40.8
M3	20.0	24.6	26.2	29.6
M4	**17.5**	**20.3**	**22.0**	**25.7**
Δ	**-2.5**	**-4.3**	**-4.2**	**-3.9**

Table 11. Image↔Image experiment (image-image translation) on CelebA-HQ and AFHQ datasets. Metrics: FID↓.

✓ *Non-text↔non-text dual learning can also benefit from structure matching.*

● Analysis

➢ Insights into Key Influencers



✓ *The dual tasks with richer structural information for the alignments will lead to better improvements.*

Thanks.