Structural Entities Extraction and Patient Indications Incorporation for Chest X-ray Report Generation

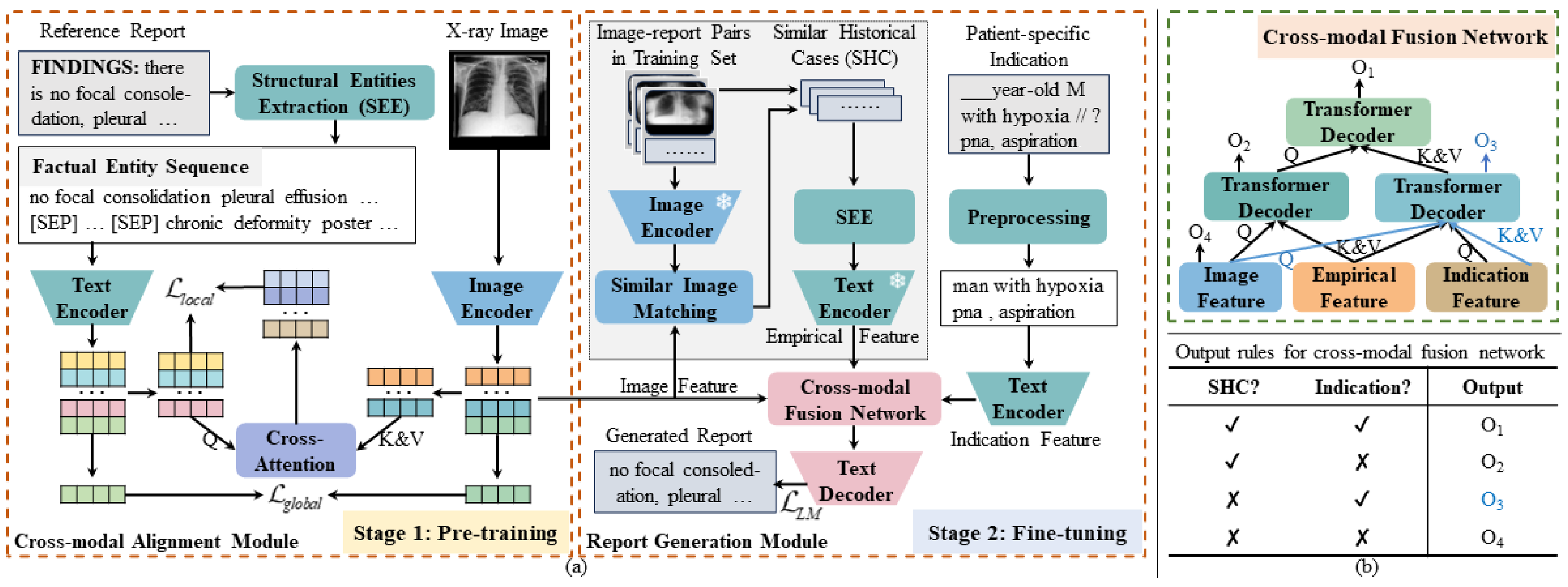
BROWN
JOHNS HOPKINS UNIVERSITY

Kang Liu, Zhuoqi Ma*, Xiaolu Kang, Zhusi Zhong, Zhicheng Jiao, Grayson Baird, Harrison Bai, Qiguang Miao*

Motivation

- **Background:** A radiology report comprises **presentation-style vocabulary**, which ensures clarity and organization, and **factual vocabulary**, which provides accurate and objective descriptions based on observable findings.
- To truly meet clinical needs, report generation processes should **incorporate patient-specific indications**, such as previous treatment history or responses to specific diagnostic requirements, which cannot be derived exclusively from medical images.
- Existing methods face challenges in effectively focusing on the cross-modal alignment between medical images and reports. This is attributed to assigning **equal weights to presentation-style elements** (e.g., sentence structure and grammar) **and factual vocabulary** (e.g., findings) in reports. Unfortunately, this limitation impacts their clinical efficacy.

Method



We propose a **two-stage** method called **S**tructural **E**ntities extraction and patient indications **I**ncorporation (SEI) for **generating chest X-ray reports**.

- ✓ **Goal:** Given a chest X-ray, our model retrieves similar historical cases in a gradient-free manner and generates a draft report based on these cases for radiologists.
- ✓ **Stage 1:** Pre-training for visual representation
  - **Structural Entities Extraction (SEE):** We remove presentation-style vocabulary from reports using RadGraph outputs and process the noise in these outputs to form factual entity sequences, which are short sentences consisting exclusively of factual vocabulary.
  - **Cross-modal Alignment:** We align medical images with **factual entity sequence** through instance-level and token-level cross-modal semantic correspondences. This approach reduces the noise in the cross-modal alignment process, facilitating gradient-free retrieval of similar historical cases from the training set.
- ✓ **Stage 2:** Fine-tuning for chest X-ray report generation
  - **Patient-specific Indications:** This field is a string **representing the patient's examination purpose or symptoms**, which may occasionally be **absent**.
  - **Cross-modal Fusion Network:** This network effectively utilizes available indications and similar historical cases through output rules, even when some samples these elements, This enables the text decoder to attend to discriminative features of X-ray images, assimilate historical diagnostic information from similar cases, and understand the examination intention of patients.

Experiments

Comparison of our SEI with SOTA approaches on the MIMIC-CXR dataset

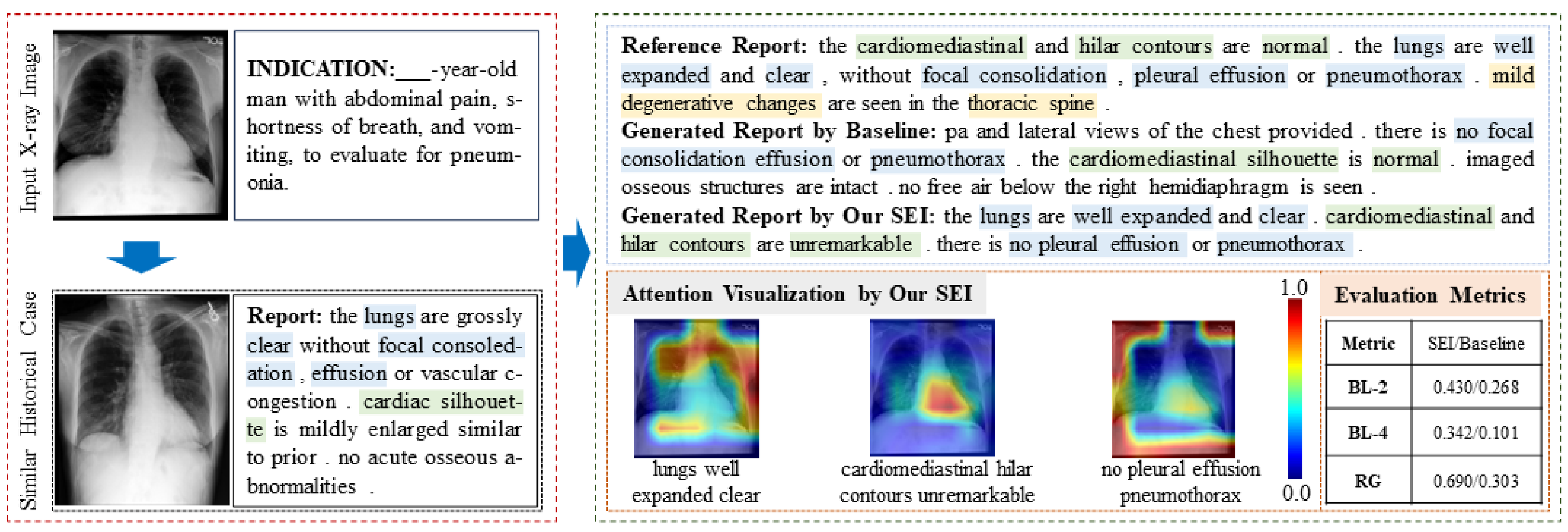
SEI-*n* indicates our SEI generates reports by referencing *n* similar historical cases

Method	M_{gt}	NLG↑ BL-2	BL-4	MTR	R_L	CE↑ RG	CX5	CX14
R2Gen [5] (EMNLP'20)	100†	0.218	0.103	0.137	0.264	0.207	0.340	0.340
	Cpl.	0.209	0.097	0.135	0.266	0.211	0.339	0.338
R2GenCMN [4] (ACL'21)	100†	0.218	0.106	0.142	0.278	0.220	0.461	0.278
	Cpl.	0.198	0.090	0.133	0.268	0.223	0.464	0.393
GSKET [29] (MedIA'22)	80†	0.228	0.115	-	0.284	-	-	0.371
CGPT2 [18] (ARTMED'23)	60†	0.248	0.127	0.155	0.286	0.223	0.463	0.391
	Cpl.	0.204	0.102	0.138	0.277	0.237	0.483	0.434
M2KT [28] (MedIA'23)	80†	0.237	0.111	0.137	0.274	0.204	0.477	0.352
	Cpl.	0.204	0.085	0.133	0.244	0.210	0.483	0.413
DCL [16] (CVPR'23)	90†	-	0.109	0.150	0.284	-	-	0.373
RGRG [20] (CVPR'23)	*Cpl.*†	**0.249**	0.126	**0.168**	0.264	-	**0.547**	0.447
SEI-0 (ours)	60	**0.268**	0.146	0.164	0.300	**0.239**	0.505	0.437
	80	0.250	0.135	0.158	**0.300**	**0.250**	0.531	0.452
	90	0.244	0.131	0.156	0.299	**0.252**	0.536	0.455
	100	0.240	0.129	0.154	0.298	**0.252**	0.539	0.457
	Cpl.	0.231	0.123	0.150	**0.297**	**0.252**	0.541	0.457
SEI-1 (ours)	60	**0.268**	**0.148**	**0.167**	**0.301**	0.236	**0.509**	**0.445**
	80	**0.257**	**0.140**	**0.162**	**0.300**	0.247	**0.535**	**0.457**
	90	**0.251**	**0.137**	**0.160**	**0.300**	0.248	**0.539**	**0.459**
	100	**0.247**	**0.135**	**0.158**	**0.299**	0.249	**0.542**	**0.460**
	Cpl.	0.238	**0.128**	0.154	0.296	0.249	0.545	**0.460**

Ablation study on MIMIC-CXR

Settings	Model	NLG↑ BL-2	BL-4	MTR	R_L	CE↑ RG	CX5	CX14
(a)	Base (R2Gen [5])	0.209	0.097	0.135	0.266	0.211	0.339	0.338
(b)	(a)+cross-modal module	0.206	0.098	0.138	0.277	0.234	0.513	0.431
(c)	SEI-1 w/o indications	0.228	0.109	0.148	0.279	0.241	0.542	**0.474**
(d)	SEI-1 w/o SHC (**SEI-0**)	0.231	0.123	0.150	**0.297**	**0.252**	0.541	0.457
(e)	**SEI-1**	**0.238**	**0.128**	**0.154**	0.296	0.249	**0.545**	0.460

Qualitative analysis on MIMIC-CXR



Conclusion:

- ✓ Using **factual entity sequences for alignment** proves to be an effective strategy.
- ✓ **Similar historical cases** provide valuable empirical insights for report generation.
- ✓ Integrating **patient-specific indications** into the report generation process significantly enhances its performance.

Welcome to our Github homepage for source codes
https://github.com/mk-runner/SEI

MK

- *Email: kangliu422@gmail.com*
- *WeChat: kangliu422*