# OBJECT-SENABTICS ALIGNED PRE-TRAINING FOR IMAGE CAPTIONING IN VIETNAMESE

Hứa Văn Sơn - 220101037

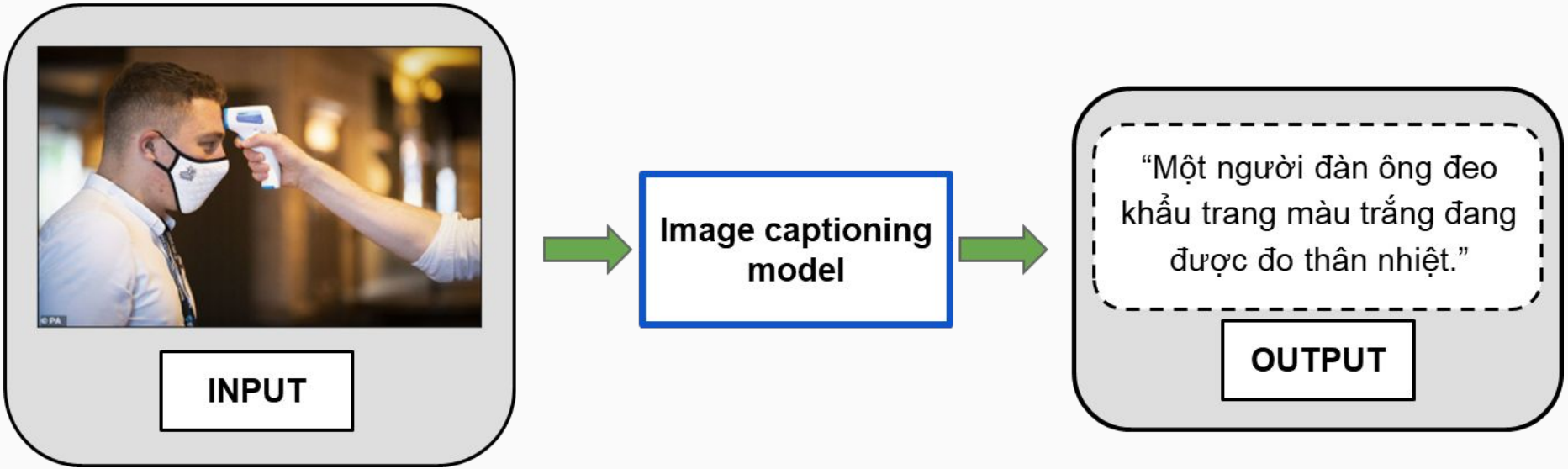
# Tóm tắt

- Lớp: CS2205.APR2023
- Link Github:
- Link YouTube video:
- Họ và Tên: Hứa Văn Sơn

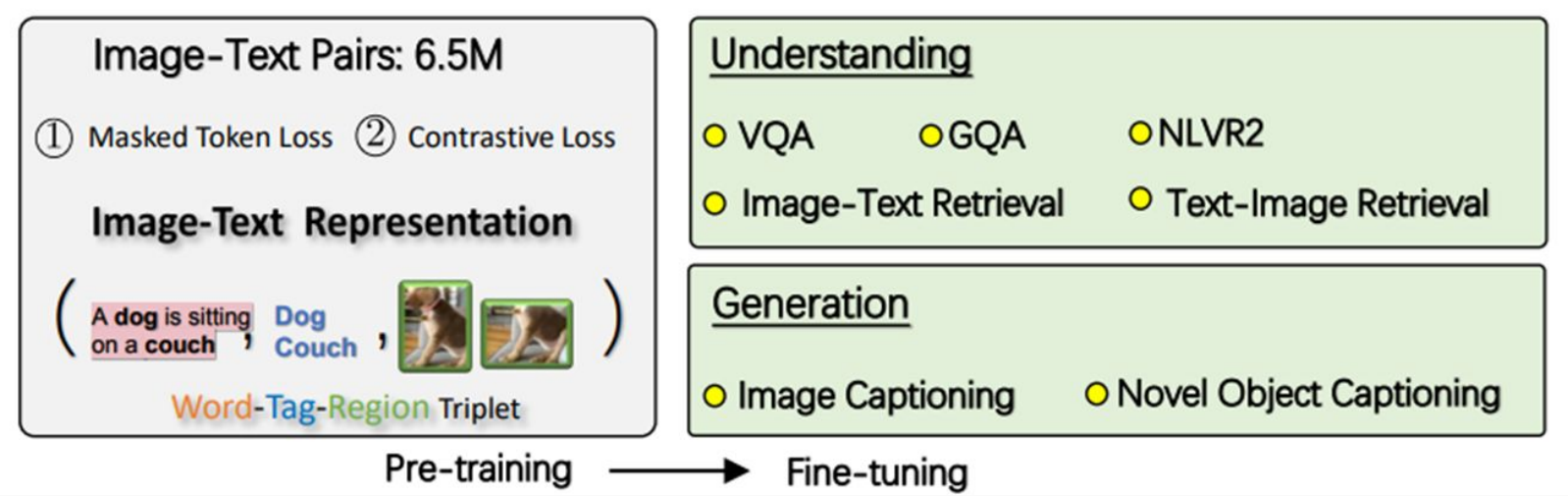
# Giới thiệu

Image Captioning in Vietnamese

- **Input**: An image
- **Output**: A caption based on content of the input image



# Giới thiệu

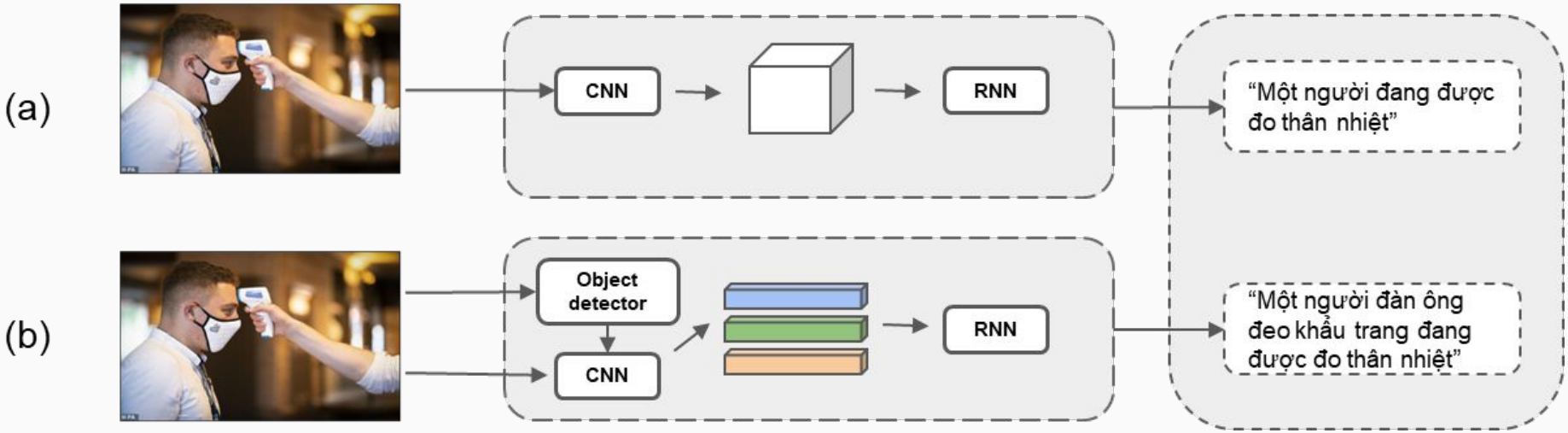



OSCAR Model

# Mục tiêu

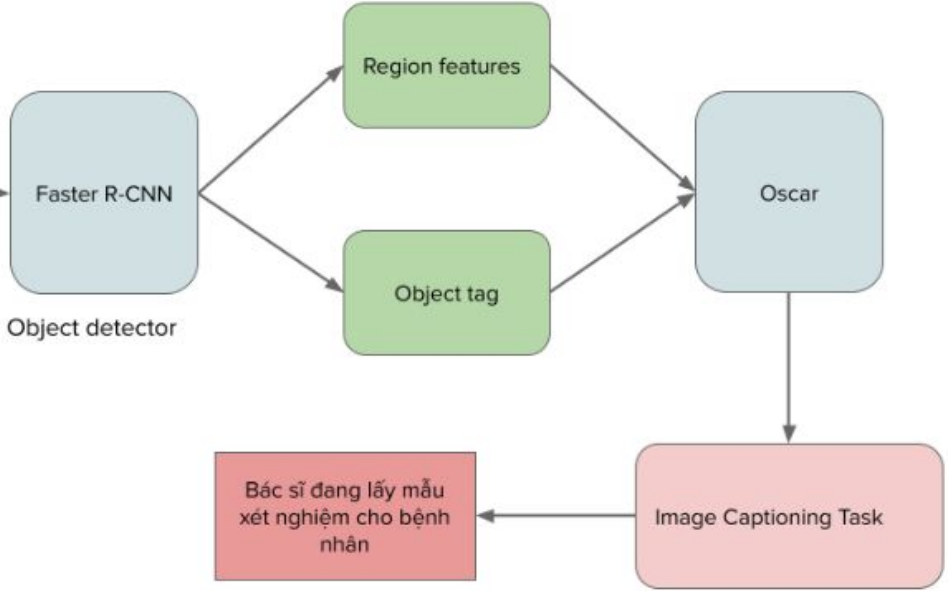
- Develop an Image Captioning model in Vietnamese
- Provide experiments and analyst insight on the effectiveness of object tags on VieCap4H dataset

# Nội dung và Phương pháp




(a) Show and tell. IEEE2015
(b) Neural baby talk. IEEE2018

# Nội dung và Phương pháp



# Kết quả dự kiến

- Successfully built a model to automatically generate text for images in Vietnamese in the healthcare field
- Using object labels increases the efficiency of text generation for the model, making the text generation more detailed and consistent with the image content. From there, opening many new research directions based on the OSCAR model, for example, it is possible to train a good object detection model for Vietnamese datasets to improve the performance and quality of the description automatic images in Vietnamese language.

# Tài liệu tham khảo

[1] Xiujun Li et al. «Oscar: Object-semantics aligned pre-training for vision- language tasks». In: European Conference on Computer Vision. Springer. 2020, pp. 121–137 (cit. on pp. i, 32, 33, 38, 41, 65).

[2] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Rus- lan Salakhudinov, Rich Zemel, and Yoshua Bengio. «Show, attend and tell: Neural image caption generation with visual attention». In: International conference on machine learning. PMLR. 2015, pp. 2048–2057 (cit. on pp. i, 2)

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. «Bert: Pre-training of deep bidirectional transformers for language understanding». In: arXiv preprint arXiv:1810.04805 (2018) (cit. on pp. 3, 29, 34, 40).

[4] Jiasen Lu, Jianwei Yang, Dhruv Batra, and Devi Parikh. «Neural baby talk». In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, pp. 7219–7228 (cit. on p. 2)