ITG-VAE

Image-conditioned Text Generation with Variational Autoencoder

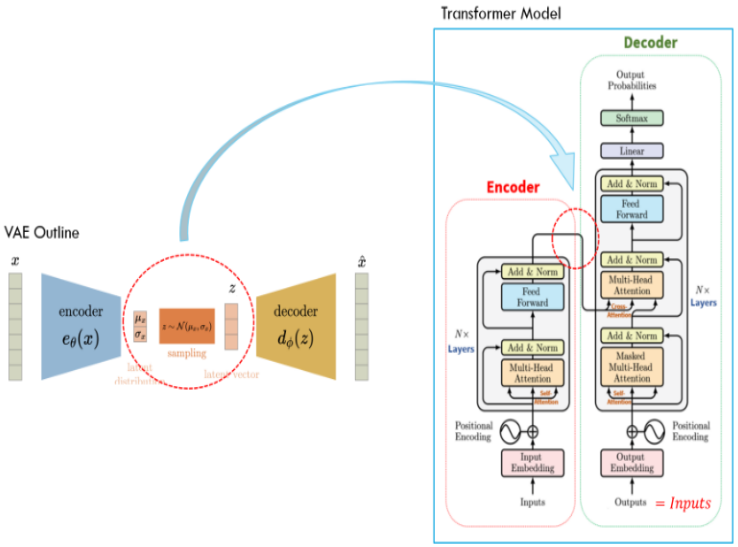
Team 20

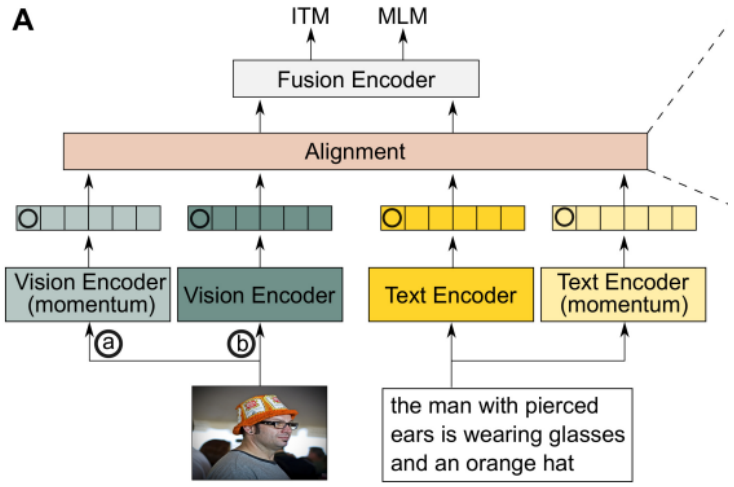
Hyeonsu Hwang
Youngwoo Shin
Kanghyeon Kim
Jihyeon Choi

Related work

- **A Transformer-Based Variational Autoencoder for Sentence Generation (IJCNN 2019)**
  - a method of combining the Variational Autoencoder and Transformer models for sentence generation.

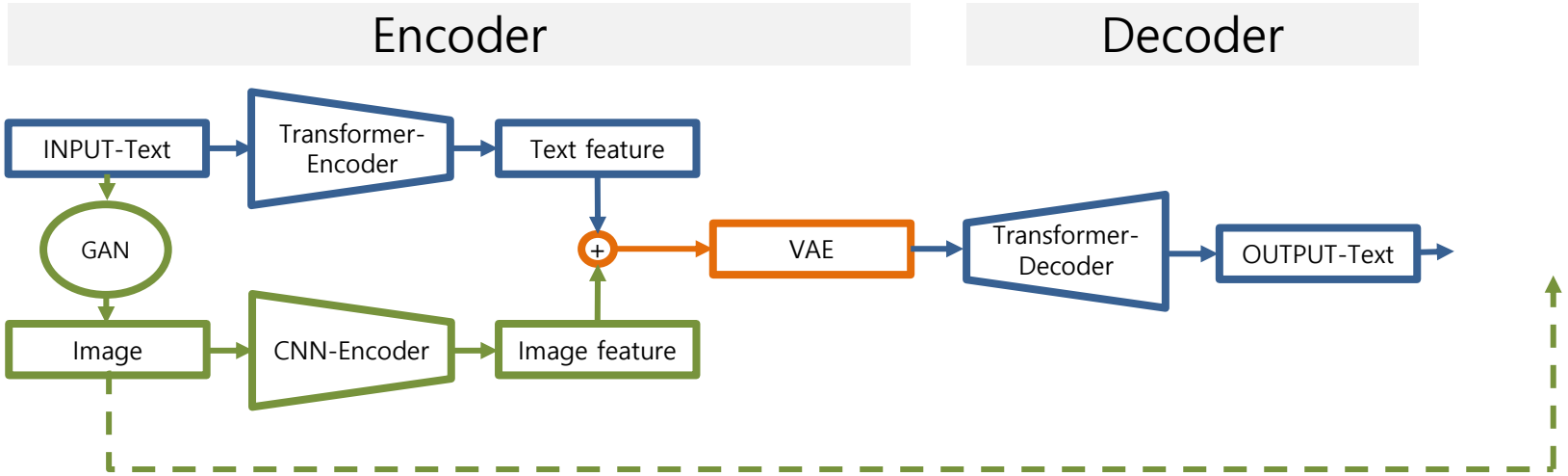
- **Vision-Language Pre-Training with Triple Contrastive Learning (CVPR 2022)**
  - a method of extracting feature vectors from images and text using a vision encoder and a text encoder, respectively, and then combining them using a fusion encoder to create multi-modal embeddings.





Proposal model

- **ITG-VAE : Image-conditioned Text Generation with Variational Autoencoder**



- We used the Penn Treebank (PTB) and the WikiText-2 datasets to train model
- **Total Loss comparison**

  (method used to compare Transformer-based VAE and LSTM-VAE in 'A Transformer-Based VAE for Sentence Generation')

  - Total Loss = KL Loss + Reconstruction Loss

  - KL Loss : calculate the difference between the distribution of latent variables and the normal distribution
  - Reconstruction Loss : calculate the difference between input and generated sentences
- **Human evaluation**

  - Evaluate the **comprehension of the text** with pairs of the generated image and the regenerated text.

감사합니다.